



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/787,283	02/26/2004	Daniel E. Tedesco	02-100B	5162

7590 12/14/2006
Dean P. Alderucci
c/o Innovation Strategies and Law Group, LLC
Box 17295
Stamford, CT 06907-9998



EXAMINER	
COUGHLAN, PETER D	
ART UNIT	PAPER NUMBER
2129	

DATE MAILED: 12/14/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 10/787,283	Applicant(s) TEDESCO ET AL.	
	Examiner Peter Coughlan	Art Unit 2129	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 07 October 2004.
- 2a) ☐ This action is FINAL. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-26 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-26 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 26 February 2004 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date <u>4/29/2004</u> | 6) <input type="checkbox"/> Other: _____ |

Detailed Action

1. Claims 1-26 are pending in this application.

Claim Rejections - 35 USC § 112

2. The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

Claims 1, 3-9, 11-13, 21, 22 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. All these claims use the term 'first user' which is not defined within the specification.

The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Art Unit: 2129

The term "substantially" in claims 1 and 5 is a relative term which renders the claim indefinite. The term "substantially" is not defined by the claim, the specification does not provide a standard for ascertaining the requisite degree, and one of ordinary skill in the art would not be reasonably apprised of the scope of the invention.

These claims have to be amended or withdrawn from consideration.

Claims 1, 19 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. These claims contain the word 'entity' and does not specify what the 'entity' is? As stated, 'entity' could be a person or a machine.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Art Unit: 2129

Claims 24, 25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Walker et al in view of Regazzoni. (U. S. Patent 6720990, referred to as Walker; 'Scanning the Issue/Technology, referred to as Regazzoni)

Claim 24

Walker teaches a step for observing a plurality of images (Walker, C1:53 through C2:4; 'receiving an image' of applicant is equivalent to 'view remote locations' of Walker.)

Walker does not teach a area in which human activity is desired to be substantially nonexistent; a step for ascertaining whether the plurality of images reliably indicates the presence of a human in the area

Regazzoni teaches of a area in which human activity is desired to be substantially nonexistent. (Regazzoni, p1361, C1:15-43; 'Human activity' of applicant is equivalent to 'people detection in highways' of Regazzoni.); a step for ascertaining whether the plurality of images reliably indicates the presence of a human in the area (Regazzoni, p1361, C1:15-43; 'Plurality of images' of applicant is equivalent to 'video surveillances' of Regazzoni.) It would have been obvious to a person having ordinary skill in the art at the time of applicant's invention to modify the teachings of Walker by specifically looking for humans as taught by Regazzoni to a area in which human activity is desired to be substantially nonexistent; a step for ascertaining whether the plurality of images reliably indicates the presence of a human in the area

Art Unit: 2129

For the purpose of filtering out only humans where humans should not be indicates an event/image needing closer inspection.

Walker teaches a step for alerting an entity based on the step for ascertaining. (Walker, C8:38-62; 'Entity to notify' of applicant is equivalent to 'the authorities' of Walker. Since the user does not know they could be one of many monitors and are under the impression they are the only one, this bypasses the 'bypasser inaction' syndrome.)

Claim 25

Walker does not teach a step for assessing an area in which human activity is desired to be substantially nonexistent.

Regazzoni teaches a step for assessing an area in which human activity is desired to be substantially nonexistent. (Regazzoni, p1361, C1:15-43; 'Human activity' of applicant is equivalent to 'people detection in highways' of Regazzoni.) It would have been obvious to a person having ordinary skill in the art at the time of applicant's invention to modify the teachings of Walker by looking specifically for human activity as taught by Regazzoni to have a step for assessing an area in which human activity is desired to be substantially nonexistent.

For the purpose of filtering out only humans where humans should not be indicates an event/image needing closer inspection.

Walker teaches a step for alerting an entity based on the step for assessing (Walker, C8:38-62; 'Entity to notify' of applicant is equivalent to 'the authorities' of

Art Unit: 2129

Walker. Since the user does not know they could be one of many monitors and are under the impression they are the only one, this bypasses the 'bypasser inaction' syndrome.)

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 1-23, 26 are rejected under 35 U.S.C. 103(a) as being unpatentable over the combination of Walker and Regazzoni, as set forth above in view of Sacchi ('A Distributed Surveillance System for Detection of Abandoned Objects in Unmanned Railroad Environments', referred to as **Sacchi**)

Claim 1

Walker teaches receiving an image from an image capture device. (Walker, C1:53 through C2:4; 'receiving an image' of applicant is equivalent to 'view remote locations' of Walker.)

Art Unit: 2129

Walker and Regazzoni do not teach in which the image capture device generates an image of an area in which human activity is desired to be substantially nonexistent. Sacchi teaches in which the image capture device generates an image of an area in which human activity is desired to be substantially nonexistent (**Sacchi**, abstract; 'Human activity is desired to be substantially nonexistent' of applicant is equivalent to 'unmanned railway environments' of Sacchi.) It would have been obvious to a person having ordinary skill in the art at the time of applicant's invention to modify the combined teachings of Walker and Regazzoni by looking for humans as taught by Sacchi to have the image capture device generates an image of an area in which human activity is desired to be substantially nonexistent.

For the purpose of filtering out only humans where humans should not be indicates an event/image needing closer inspection.

Walker teaches determining information related to the area (**Walker**, C1:28-36; 'Determining information' of applicant is equivalent to 'view customer behavior' of Walker.); receiving a request for a first user to monitor(**Walker**, C5:48-67; 'Request for a first user to monitor' of applicant is equivalent to 'user first request to monitor' of Walker.); receiving a user identifier(**Walker**, C5:48-67; 'User identifier' of applicant is equivalent to 'record of the user' of Walker.); verifying that the user identifier corresponds to the first user(**Walker**, C5:48-67; 'Verifying' of applicant is equivalent to 'log on the central server. '); providing the first user with the image. (**Walker**, C1:53 through C2:4; 'providing an image' of applicant is equivalent to 'view remote locations' of Walker.)

Art Unit: 2129

Walker does not teach receiving a response to the image by the first user, in which the response comprises an indication that a human is present in the image.

Regazzoni teaches receiving a response to the image by the first user, in which the response comprises an indication that a human is present in the image.

(Regazzoni, p1361, C1:15-43; 'Human is present' of applicant is equivalent to 'people detection in highways' of Regazzoni.) It would have been obvious to a person having ordinary skill in the art at the time of applicant's invention to modify the teachings of Walker by looking for humans as taught by Regazzoni to receiving a response to the image by the first user, in which the response comprises an indication that a human is present in the image.

For the purpose of filtering out only humans where humans should not be indicates an event/image needing closer inspection.

Walker teaches providing additional users with the image (Walker, C3:46-59 and C4:35-57; 'Additional users' of applicant is equivalent to 'user devices 300a-c' of Walker.); receiving responses to the image by the additional users (Walker, C4:35-57; 'Responses' of applicant is equivalent to 'responses' of Walker.); evaluating the received responses (Walker, C9:61 through C10:16; 'Evaluating the responses' of applicant is equivalent to 'evaluates the responses' of Walker.); determining, based on the information related to the area, an entity to notify (Walker, C8:38-62; 'Entity to notify' of applicant is equivalent to 'the authorities' of Walker.); and notifying the entity. (Walker, C8:38-62; Since the user does not know they could be one of many monitors

Art Unit: 2129

and are under the impression they are the only one, this bypasses the 'bypasser inaction' syndrome.)

Claim 2

Walker teaches receiving a unique identifier from the image capture device(Walker, C3:46-59; 'User devices 300a-c' are connected to a web-based service. Therefore each user device has to have its own unique IP address. Therefore, 'unique identifier' of applicant is equivalent to each unique IP address of user's devices.); accessing a record in a database using the unique identifier(Walker, C4:7-20; 'Accessing a record' of applicant is equivalent to accessing the server by using the IP address of the server.); and determining, from the record, contact information for the area. (Walker, C4:7-20; 'Determining from the record, contact information' of applicant is equivalent to 'registering' of Walker.)

Claim 3

Walker teaches transmitting the image to an internet protocol address which is based on the first user.(Walker, C2:5-17; Images stored on a server which are part of a web based system have a IP address.)

Claim 4

Art Unit: 2129

Walker teaches posting the image on a Web site. (Walker, C2:5-17; Walker discloses a web based system, thus images stored in a server are 'posted' on a web site.)

Claim 5

Walker teaches receiving a request for a first user to monitor (Walker, C5:48-67; 'Request for a first user to monitor' of applicant is equivalent to 'user first request to monitor' of Walker.); verifying the first user. (Walker, C5:48-67; 'Verifying' of applicant is equivalent to 'log on the central server.')

Walker and Regazzoni do not teach providing the first user with an image of an area in which human activity is desired to be substantially nonexistent.

Sacchi teaches providing the first user with an image of an area in which human activity is desired to be substantially nonexistent. (Sacchi, abstract; 'Human activity is desired to be substantially nonexistent' of applicant is equivalent to 'unmanned railway environments' of Sacchi.) It would have been obvious to a person having ordinary skill in the art at the time of applicant's invention to modify the combined teachings of Walker and Regazzoni by looking for human activity as taught by Sacchi to providing the first user with an image of an area in which human activity is desired to be substantially nonexistent.

For the purpose of filtering out only humans where humans should not be indicates an event/image needing closer inspection.

Art Unit: 2129

Walker teaches receiving a response to the image by the first user(**Walker**, C4:35-57; 'Responses' of applicant is equivalent to 'responses' of Walker.); and evaluating the received response. (**Walker**, C9:61 through C10:16; 'Evaluating the responses' of applicant is equivalent to 'evaluates the responses' of Walker.)

Claim 6

Walker teaches receiving an identifier(**Walker**, C5:48-67; 'User identifier' of applicant is equivalent to 'record of the user' of Walker.); and determining that the identifier identifies a prior registration. (**Walker**, C4:7-20; 'Identifier identifies a prior registration' of applicant is equivalent to 'after registering, users can simply present their user identifier to the central server.)

Claim 7

Walker teaches determining an attentiveness of the first user. (**Walker**, abstract; 'determining an attentiveness' of applicant is equivalent to 'measuring user attentiveness' of Walker.)

Claim 8

Walker teaches requesting that the first user respond to a false positive. (**Walker**, C4:35-57; Responding to a false positive of applicant is equivalent to 'test the guards attentiveness' of Walker.)

Art Unit: 2129

Claim 9

Walker teaches providing the first user with a false positive image(Walker, C4:35-57; Providing false positive image of applicant is equivalent to 'transmitting test communication' of Walker.); receiving a response to the false positive image by the first user. (Walker, C4:35-57; 'Receiving a response' of applicant is equivalent to 'responds to test communication' of Walker.)

Claim 10

Walker does not teach determining whether the response to the false positive image indicates that a human is present in the image.

Regazzoni teaches determining whether the response to the false positive image indicates that a human is present in the image. (Regazzoni, p1361, C1:15-43; 'Human is present' of applicant is equivalent to 'people detection in highways' of Regazzoni.) It would have been obvious to a person having ordinary skill in the art at the time of applicant's invention to modify the teachings of Walker by testing the user with human images where humans should not be as taught by Regazzoni to determining whether the response to the false positive image indicates that a human is present in the image.

For the purpose of testing the user which determines the user's rating.

Claim 11

Walker teaches determining a reaction time of the first user. (Walker, C6:39-56; 'Reaction time' of applicant is equivalent to 'response time' of Walker.)

Claim 12

Walker teaches selecting the first user from a plurality of users. (Walker, C5:48-67; 'Selecting the first user' of applicant is equivalent to 'user first request' of Walker.)

Claim 13

Walker teaches selecting the first user from the plurality of users based on the image. (Walker, C4:7-20; Selecting a user based on the image of applicant means selecting a user based on their rating of attentiveness on testing. This is disclosed in Walker by requiring a minimum user rating.)

Claim 14

Walker teaches providing at least one additional user with the image. (Walker, C4:35-57; 'One additional user' of applicant is equivalent to 'plurality of users' of Walker.)

Claim 15

Walker teaches providing at least one additional user with the image is performed based on the response to the image. (Walker, C4:7-20, C4:35-57; Providing the additional user with an image of applicant is equivalent to 'monitored by a plurality of users' of Walker. 'Performance based' of applicant is equivalent to 'attentiveness' of Walker.)

Claim 16

Walker teaches determining, based on the response to the image, a number (Walker, C4:35-57; 'A number' of applicant is equivalent to 'user's rating' of Walker.); selecting a plurality of additional user, in which the cardinality of the plurality is at least the number (Walker, C4:7-20; 'Additional users' with cardinality of applicant is equivalent to users with minimum rating of Walker.); and 1providing the plurality of additional users with the image. (Walker, C4:35-57; Providing additional users with the image of applicant is equivalent to 'monitored by a plurality of users' of Walker.)

Claim 17

Walker teaches determining a response time in receiving the response to the image. (Walker, C6:39-56; 'Determining a response time' of applicant is equivalent to testing for a 'response time' of Walker.)

Claim 18

Walker does not teach in which the response is one of: an indication that a human is present in the image, an indication that no human is present in the image, and an indication of uncertainty whether a human is present in the image.

Regazzoni teaches in which the response is one of: an indication that a human is present in the image, an indication that no human is present in the image (Regazzoni, p1361, C1:15-43; Regazzoni discloses if a human is or is not present in highways.), and

Art Unit: 2129

an indication of uncertainty whether a human is present in the image. (**Regazzoni**, p1360, C1:16 through C2:25; 'Indication of uncertainty' of applicant is equivalent to 'error rate of less than 1%' of Regazzoni.) It would have been obvious to a person having ordinary skill in the art at the time of applicant's invention to modify the teachings of Walker by having one of three possible outcomes regarding 'human' images as taught by Regazzoni to which the response is one of: an indication that a human is present in the image, an indication that no human is present in the image, and an indication of uncertainty whether a human is present in the image.

For the purpose of knowing there is or is not a human present along with a threshold application for possible outcomes

Claim 19

Walker teaches determining, based on the received response, whether to notify an entity. (**Walker**, C11:38-64; 'Whether to notify an entity' of applicant is equivalent to 'determined whether the reported emergency is legitimate' of Walker.)

Claim 20

Walker teaches initiating a telephone call to a predetermined telephone number. (**Walker**, C11:38-64; 'Initiating a telephone call' of applicant is equivalent to 'communicates to the user a phone number' of Walker.)

Claim 21

Art Unit: 2129

Walker teaches adjusting a rating of the first user based on the received response. (Walker, C11:38-64; 'Adjusting a rating' of applicant is equivalent to 'lowers the rating in the user database' of Walker.)

Claim 22

Walker teaches compensating the first user. (Walker, C10:17-46; 'Compensating' of applicant is equivalent to 'pay' of Walker.)

Claim 23

Walker teaches compensating the first user based on the received response. (Walker, C10:17-46; 'Compensating' of applicant is equivalent to 'pay' of Walker. 'Received responses' of applicant is directly related to 'higher crime rates' of Walker.)

Claim 26

Walker teaches means for receiving images. (Walker, C1:53 through C2:4; 'receiving an image' of applicant is equivalent to 'view remote locations' of Walker.) Walker does not teach an area in which human activity is desired to be substantially nonexistent.

Regazzoni teaches an area in which human activity is desired to be substantially nonexistent. (Regazzoni, p1361, C1:15-43; 'Human activity' of applicant is equivalent to 'people detection in highways' of Regazzoni.) It would have been obvious to a person having ordinary skill in the art at the time of applicant's invention to modify the combined

Art Unit: 2129

teachings of Walker by looking for human activity as taught by Regazzoni to have an area in which human activity is desired to be substantially nonexistent.

For the purpose of filtering out only humans where humans should not be indicates an event/image needing closer inspection.

Walker teaches means for distributing the images for at least partial analysis. (Walker, abstract; 'Means for distributing the images' of applicant is accomplished by the 'server' of Walker.)

Walker and Regazzoni do not teach means for calculating an analysis of the images.

Sacchi teaches means for calculating an analysis of the images. (Sacchi, abstract; 'Analysis of the image' of applicant is equivalent to 'image processing system' of Sacchi.) It would have been obvious to a person having ordinary skill in the art at the time of applicant's invention to modify the combined teachings of Walker and Regazzoni by using an algorithm for image analysis as taught by Sacchi to have means for calculating an analysis of the images.

For the purpose of using a method for finding items within an image that might need further review.

Walker teaches means for warning an entity based on the analysis. (Walker, C8:38-62; 'Entity to notify' of applicant is equivalent to 'the authorities' of Walker. Since the user does not know they could be one of many monitors and are under the impression they are the only one, this bypasses the 'bypasser inaction' syndrome.)

Conclusion

4. The prior art of record and not relied upon is considered pertinent to the applicant's disclosure.

-U. S. Patent 6476858: Ramirez Diaz

-U. S. Patent 6271752: Vaio

-U. S. Patent 6166729: Acosta

-U. S. Patent 5909548: Klein

-U. S. Patent 5857190: Brown

-U. S. Patent 5794210: Goldhaber

-U. S. Patent 5786746: Lombardo

-U. S. Patent 5759101: Von Kohorn

-U. S. Patent 5412708: Katz

-U. S. Patent 5034807: Von Kohorn

-U. S. Patent 4622538: Whynacht

-U. S. Patent 4511886: Rodriguez

Art Unit: 2129

5. Claims 1-26 are rejected.

Correspondence Information

6. Any inquiry concerning this information or related to the subject disclosure should be directed to the Examiner Peter Coughlan, whose telephone number is (571) 272-5990. The Examiner can be reached on Monday through Friday from 7:15 a.m. to 3:45 p.m.

If attempts to reach the Examiner by telephone are unsuccessful, the Examiner's supervisor David Vincent can be reached at (571) 272-3687. Any response to this office action should be mailed to:

Commissioner of Patents and Trademarks,
Washington, D. C. 20231;

Hand delivered to:

Receptionist,
Customer Service Window,
Randolph Building,
401 Dulany Street,
Alexandria, Virginia 22313,
(located on the first floor of the south side of the Randolph Building);

Art Unit: 2129

or faxed to:


(571) 273-8300 (for formal communications intended for entry.)

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have any questions on access to Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll free).

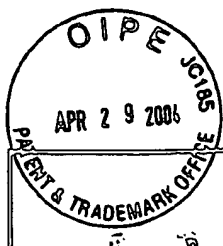


Peter Coughlan

12/08/2006



DAVID VINCENT
SUPERVISORY PATENT EXAMINER



INFORMATION DISCLOSURE CITATION

Docket Number

02-100B

Serial Number

10/787,283

Applicants

TEDESCO et al.

Filing Date

February 26, 2004

Group Art Unit 2129

~~Not Yet Assigned~~

U.S. PATENT DOCUMENTS

EXAMINER INITIAL	REF	DOCUMENT NUMBER	DATE	NAME	CLASS	SUB- CLASS	FILING DATE IF APPROPRIAT E
PC	A	4,458,266	07/03/84	Mahoney	358	105	
	B	4,511,886	04/16/85	Rodriguez	340	534	
	C	4,622,538	11/11/86	Whynacht et al.	340	506	
	D	4,646,145	02/24/87	Percy et al.	358	84	
	E	4,861,154	08/29/89	Sherwin et al.	351	205	
	F	4,955,388	09/11/90	Silberstein	128	731	
	G	4,982,346	01/01/91	Girouard et al.	364	550	
	H	5,247,433	09/21/93	Kitaura et al.	364	188	
	I	5,412,708	05/02/95	Katz	348	14	
	J	5,519,669	05/21/96	Ross et al.	367	93	
	K	5,550,581	08/27/96	Zhou	348	17	
✓	L	5,571,057	11/05/96	Ayers	463	36	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NUMBER	DATE	COUNTRY	CLASS	SUB- CLASS	Translation	
							Yes	No

OTHER DOCUMENTS (Including author, Title, Date, Pertinent Pages, Etc.)

--	--	--

EXAMINER

/Peter Coughlan/

DATE CONSIDERED:

12/06/2006

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

INFORMATION DISCLOSURE CITATION	Docket Number 02-100B	Serial Number 10/787,283
	Applicants TEDESCO et al.	
	Filing Date February 26, 2004	Group Art Unit 2129 Not Yet Assigned

U.S. PATENT DOCUMENTS

EXAMINER INITIAL	REF	DOCUMENT NUMBER	DATE	NAME	CLASS	SUB-CLASS	FILING DATE IF APPROPRIATE
PC	M	5,598,456	01/28/97	Feinberg	379	42	
	N	5,734,335	03/31/98	Brogi et al.	340	870.05	
	O	5,786,746	07/28/98	Lombardo et al.	340	286.07	
	P	5,794,210	08/11/98	Goldhaber et al.	705	14	
	Q	5,857,190	01/05/99	Brown	707	10	
	R	5,909,548	06/01/99	Klein et al.	395	200.47	
	S	5,948,054	09/07/99	Nielsen	709	200	
	T	US 6,271,752 B1	08/07/01	Vaios	340	541	
	U	US 6,297,825 B1	10/02/01	Madden et al.	345	419	
	V	2002/0087413 A1	07/04/02	Mahaffy et al.	705	16	
	W	US 6,476,858 B1	11/05/02	Ramirez Diaz et al.	348	159	
↓	X	US 6,538,689 B1	03/25/03	Chang	348	159	

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NUMBER	DATE	COUNTRY	CLASS	SUB-CLASS	Translation	
							Yes	No
PC	Y	WO 97/22074	06/19/97	PCT				

OTHER DOCUMENTS (Including author, Title, Date, Pertinent Pages, Etc.)

EXAMINER /Peter Coughlan/	DATE CONSIDERED: 12/06/2006
EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.	

INFORMATION DISCLOSURE CITATION		Docket Number 02-100B	Serial Number 10/787,283
		Applicants TEDESCO et al.	
		Filing Date February 26, 2004	Group Art Unit 2129 Not Yet Assigned
OTHER DOCUMENTS (Including author, Title, Date, Pertinent Pages, Etc.)			
PC	Z	Risen, James, "Rise in Temp Workers Leads to Sharp Fall in Job Security", The Tech, February 15, 1994, Vol. 114, No. 6, pg. 2.	
	AA	Edmondson, Gail et al., "A Tidal Wave of Temps (int'l edition)", Business Week, Copyright 1997 by the McGraw-Hill Companies, Inc.	
	BB	Spera, Vincent, "The Low Wage Labor Market: Problems and Solutions", American Youth Policy Forum, September 5, 1997, Capitol Hill.	
	CC	Bott, Jennifer, "Eyes on Road, Hands on the Wheel? Researchers Want to See Whether Technical Gadgets and Devices are Creating Sensory Overloads for Drivers.", Akron Beacon Journal, May 25, 1998, Section: Business, pg. D1.	
	DD	Website: "Remote Video Monitoring Software by PriVIDEeye", (http //www privid-eye com/index html), download date: August 5, 1998.	
	EE	Website: "Remote video surveillance and monitoring of high-tech, PC-cam adapter!", (http //www novexcanada com/), download date: August 5, 1998.	
	FF	Website: "Spies: Surveillance, Security, Wiretaps, Prot...", (http //www spyzone com/CCSSL1 html), download date: August 5, 1998.	
	GG	Website: "SpyZone Com: CCS Group: Company Profile", (http //www spyzone com/CCSSZ1 html), download date: August 5, 1998.	
	HH	Gaskin, James E., "Cyberspace goes blue-collar", Inter@ctive Week, August 19, 1998.	
	II	Ruppe, David, "ABCNEWS.com: Working Poor Reap Small Gains in New Economy", (http //abcnews go com/sections/business/DailyNews/poverty_000628 html), Copyright 2000 ABC News Internet Ventures.	
	JJ	Website: "FDIC Consumer News – Winter 2001/2002 – Weathering a Financial Storm", (http //www fdic gov/consumers/consumer/news/cnwin0102/cvrstry html).	
✓	KK	Holzer, Harry J. et al., "Job Change and Job Stability Among Less-Skilled Young Workers", The Harris School - University of Chicago, Working Paper Series, Copyright 2002-2003.	
EXAMINER		/Peter Coughlan/	DATE CONSIDERED: 12/06/2006
EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.			

INFORMATION DISCLOSURE CITATION		Docket Number	Serial Number
		02-100B	10/787,283
		Applicants	
		TEDESCO et al.	
		Filing Date	Group Art Unit
		February 26, 2004	2129 Not Yet Assigned
OTHER DOCUMENTS (Including author, Title, Date, Pertinent Pages, Etc.)			
PC	LL	Website: "Even With Steady Job, You're a 'Temp'", PRWeb, April 13, 2002, (http //www prweb com/releases/2002/4/prweb36788 php).	
	MM	Jacobson, Louis, "Science: Security Experts Picking Up Good Vibrations", The Washington Post, December 30, 2002, pg. A9.	
	NN	Erickson, Kurt, "Big Debate on Minimum Wage", The Pantagraph, February , 23, 2003, (http //www livingwagecampaign org/pc php?p=2026).	
	OO	Website: "Institute on Education and the Economy (Reorganization of Workplace – Project 3)", (http www tc Columbia edu/~ice/Workpl3 htm), download date: March 11, 2003.	
	PP	Holzer, Harry J., and Lalonde, Robert J., "Job Instability Among Young Adults", Research Summary, The Irving B. Harris Graduate School of Public Policy Studies, The University of Chicago, (http www harrisschool uchicago edu/publications/rs/1-3 html), download date: March 11, 2003	
	QQ	Website: "American Consumer Opinion – Paid Surveys FAQ", (http //www acop com/FAQ asp), download date: March 13, 2003.	
	RR	Website: "Earn money – online surveys – get paid for your opinion", (http www tamingthebeast net/misc/online-surveys htm), download date: March 13, 2003.	
↓	SS	Registration Confirmation email: "Congratulations! You have completed the registration process and are now a member of American Consumer Opinion TM !", American Consumer Opinion, dated: April 13, 2003.	
EXAMINER /Peter Coughlan/		DATE CONSIDERED: 12/06/2006	
EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.			

Notice of References Cited

Application/Control No.

10/787,283

Applicant(s)/Patent Under
Reexamination
TEDESCO ET AL.

Examiner

Peter Coughlan

Art Unit

2129

Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
*	A	US-6,720,990	04-2004	Walker et al.	348/143
	B	US-			
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	'A Distributed Surveillance System for Detection of Abandoned Objects in Unmanned Railway Environments': Sacchi, Regazzoni, 2000, IEEE, IEEE transactions on vehicular technology, Vol, 49, No. 5, pages 2013-2026
	V	'Scanning the Issue/Technology': Regazzoni, 2001, IEEE, 0018-9219(01)08438-9, pages 1355-1367
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

A Distributed Surveillance System for Detection of Abandoned Objects in Unmanned Railway Environments

Claudio Sacchi and Carlo S. Regazzoni, *Senior Member, IEEE*

Abstract—In this paper, a distributed video-surveillance system for the detection of dangerous situations related to the presence of abandoned objects in the waiting rooms of unattended railway stations is presented. The image sequences, acquired with a monochromatic camera placed in each guarded room, are processed by a local PC-based image-processing system, devoted to detecting the presence of abandoned objects. When an abandoned object is recognized, an alarm issue is transmitted to a remote control center, located few miles far from the guarded stations. A multimedia communication system based on direct sequence code-division multiple-access (DS/CDMA) techniques aims at ensuring secure and noise-robust wireless transmission links between the guarded stations and the remote control center, where the processing results are displayed to the human operator. Results concern: 1) the performances of each local image processing system in terms of false-alarm and misdetection probabilities, and 2) the performances of the CDMA multimedia transmission system in terms of bit error rates (BERs) and quality of service (QoS).

Index Terms—Image processing, multimedia communication, rail transportation, site security monitoring, surveillance.

I. INTRODUCTION

THE increasing request for security and efficiency in the field of public transportation systems, for both people and goods, has resulted in a corresponding increasing interest in the use of the most advanced video-based surveillance techniques in order to provide an automatic continuous monitoring of roads, railways, vehicles, and land transport infrastructures (e.g., railway stations, highway toll-gates, etc.). The main objectives of a surveillance system in transport environments concern the detection and the prevention of dangerous situations, e.g., vehicle accidents, run-over pedestrians, people falling over railway tracks, cars that stopped at unattended level crossings, etc., and the management of the vehicular traffic, in order to optimize the flow on roads and highways. Several applications of image processing and advanced data transmission techniques to the surveillance of transport environments have been presented in the literature [1]–[10].

Manuscript received January 4, 1999; revised February 15, 2000. This work was supported in part by the National Research Council (CNR) of Italy within the framework of the "National Targeted Project Transport 2" (PFT2) Research Programme, and by the Italian Ministry of University and Scientific Research (MURST) under the National Interest Scientific Research Programme.

The authors are with the Department of Biophysical and Electronic Engineering (DIBE), University of Genoa, 16145, Genoa, Italy (e-mail: carlo@dibe.unige.it).

Publisher Item Identifier S 0018-9545(00)07906-8.

Concerning road transport, the *AUTOSCOPE* system, developed in the USA in the mid-1980s [1], is one of the best-known examples of video-based highway traffic monitoring systems. Image sequences acquired with a camera are processed by a microprocessor system that detects in real time the presence or the passing of a vehicle in the camera field of view. Another system for real-time accident prevention and traffic monitoring has been developed in Europe and is described in [2]. The system, known as *TRISTAR*, processes images coming out from cameras placed near highway lanes and produces alarm signals when a potentially dangerous situation (e.g., accident risk) is detected. A further example of a video-based system for traffic monitoring and management is presented in [3]. The described system can be addressed to obtain a visual tracking modality (i.e., vehicle tracking and pedestrian tracking) for a traffic advisory system. In [3] it is shown that the exploitation of advanced image-processing techniques for moving-object detection and tracking can be a valid support to increase the margin of safety in a large variety of common traffic situations.

A quite futuristic, though very interesting, research field is that of the video-based control procedures for computer-driven unmanned vehicles. In [4], a video-camera based method for determining the location and the rotation of autonomous vehicles is proposed. In [5] the development of a portable hardware/software neural-network module for autonomous vehicle following is described. An autonomous vehicle following is defined as a vehicle controlling its own steering and speed by following a lead vehicle [5]. A neural-network approach is exploited to determine the nonlinear relation between the observed range and heading angle and the controllable steering-wheel angle and speed. The data on the range and the heading angle are acquired by a stereo-vision system, and a neural-network-based image-processing system generates the driving command as its own issues. The synergies between vehicle recognition and tracking processes for autonomous vehicle driving are studied in [6]. Object recognition is performed in order to focus attention on interesting parts of a guarded scene and to assign symbolic meanings to them. Tracking is used to maintain a correspondence between the objects identified at successive recognition instants.

In railway environments, traffic and car safety management tasks, such as: headway between trains, speed regulation, and collision avoidance, are generally implemented by railway signaling systems, using secure and noise-robust digital radio transmission techniques. For this reason, video-surveillance applications in railway transport essentially aim at meeting the request for a protection against accidental or intentional

situations that may risk the safety of passengers. This request has become particularly urgent in metropolitan railway environments and in general in urban railway lines. In this sense, the European *CROMATICA* (*Crowd Management with Telematic Imaging and Communication Assistance*) project [7] is addressed to measure continuously the crowd flow at metro stations in order to detect the conditions of "abnormal crowd" (e.g., overcrowding, unexpected patterns of motion, queues) and prevent dangerous situations related to falls on tracks, vandal acts, personal attacks, etc., which might cause serious problems to a large number of passengers. Another European project aimed at enforcing the safety of urban railway transport is *AVS-PV* (*Advanced Video Surveillance-Prevention of Vandalism in the Metro*). The main objective of this project [8] is to detect behaviors that are typical for potential vandals in metro stations. The AVS-PV image processing system is devoted to pointing out some particular "strange" behaviors, such as a single person remaining for abnormally long time at the same place without taking a train, a "gang behavior," i.e., a number of persons belonging to a group, but not forming a group from a visual point of view, and the "agitated" behavior of a single person or of a small group of persons.

The problem of the remote monitoring of unmanned level-crossings to detect intruder objects (e.g., cars not moving on the tracks) is considered in [9] and [10]. The works in [9] and [10] describe a very low bit-rate image coding system for the transmission of the shapes of detected intruder objects (e.g., cars) to a remote operator.

Our paper aims at describing a distributed video-surveillance system for the detection of abandoned objects in unattended railway stations, a surveillance task that is strictly required for the safety of the urban transport users. Most of the situations involving the presence of abandoned objects in waiting rooms, (e.g., bags, parcels, etc.), are caused with careless passengers. However, some well-known recent terroristic attacks have pointed out that a small abandoned object can hide a highly destructive bomb. For this reason, it is reasonable to exploit the most advanced image-processing and digital-communication techniques, to detect such potentially dangerous situations and to transmit the corresponding alert signals to the security police. The proposed system acquires multimedia information (i.e., image sequences) from monochromatic cameras placed in the guarded waiting rooms. The information is then processed by a *local PC-based image processing system*, which is devoted to detecting in real time the variations occurring in a guarded scene with respect to a background scene, which represents the waiting room of an unattended railway station without any extraneous object, and to assigning each variation to a precise *class*, belonging to a limited set. When a change in the monitored scene is classified as an *abandoned object*, an alarm issue is generated by the local processing system and the corresponding alert information is transmitted to a remote control center by a wireless digital radio equipment, using *direct-sequence spread-spectrum* (DS/SS) techniques [19], for a secure and noise-robust link between the local and remote processing sites. The multiple access protocol is based on the application of *asynchronous code-division multiple-access* (DS/CDMA) techniques [19], which can provide the best

efficiency in terms of bandwidth occupation and simplicity of implementation. The remote control center should be some miles far from the unattended guarded stations (e.g., in the premises of a station located in a central metropolitan area). Remote processing is devoted to presenting to human operators the alert information transmitted by local processing systems, when an abandoned object is effectively detected. The transmission of alert information should be exhaustive and not redundant, in order to make intelligible the alarm issue to the human operator without a great expense in terms of bandwidth occupation. The economic usage of the band is a key problem in remote advanced video surveillance (AVS) applications. Such applications require the real-time transmission of large amount of data over the uplink (i.e., base-station) channel, whose bandwidth is generally less wide than the one available for the downlink (i.e., station-base) direction [18]. An accurate selection and the proposed combined source and channel coding of the information to be transmitted to the remote control center have been studied in order to meet the above-mentioned uplink bandwidth constraints. The paper is organized as follows. Section II provides a global description of the proposed system; Section III describes a local image processing system for detecting abandoned objects. Section IV deals with the DS/CDMA multimedia communication system; Section V details image processing at the remote control center; Section VI reports some numerical results together with some considerations about the use of color image sequences for abandoned object detection (instead of the monochromatic ones considered in the proposed analysis); finally in Section VII conclusions are drawn.

II. SYSTEM OVERVIEW

The global architecture of the proposed remote video-surveillance system is shown in Fig. 1.

In each unattended station, a local processing of the image sequences acquired with a monochromatic camera installed in the waiting room is performed to detect dangerous situations related to the presence of abandoned objects. The local processing and communication system is presented in Fig. 2.

The monochromatic camera acquires image sequences from the guarded waiting room. Fig. 2 depicts the interior of a waiting room, where a wooden bench is the background scene, whereas the camera, the suit-case and the folder are to be considered as abandoned objects. The image sequences are digitized by an acquisition board, which can be installed inside the PC. The software algorithms for abandoned object detection process in real time the digitized sequences and the processing results are then transmitted to the remote control center. The transmission system first transmits the background image of the guarded environment. The background image can be periodically refreshed, when a significant change in the scene occurs (e.g., a noticeable variation in light, or a change in the inner configuration). In this case, the local processing system provides the transmission of a new background. When an abandoned object is recognized, the transmission of a determinate multimedia alert information to the remote control center is automatically activated. In the next sections, the local

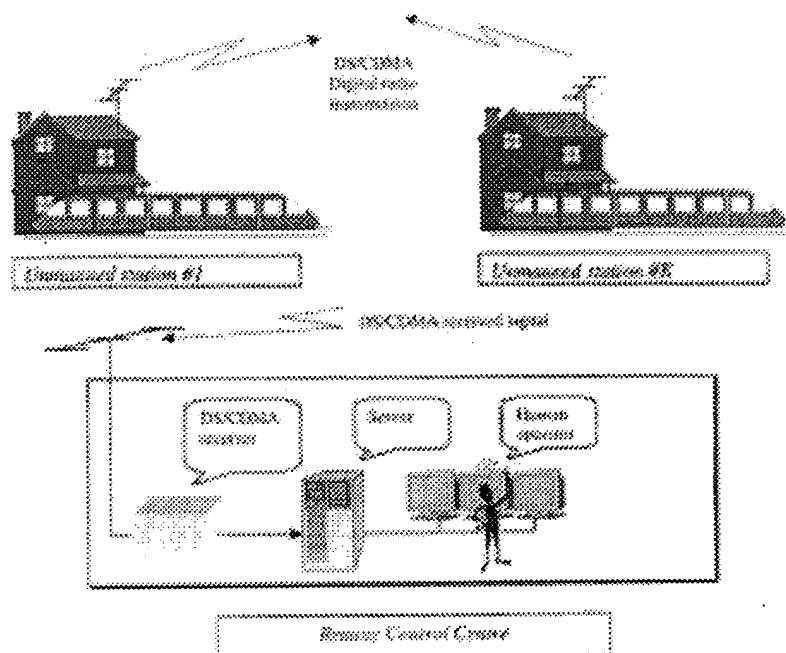


Fig. 1. Global scheme of the abandoned object detection system.

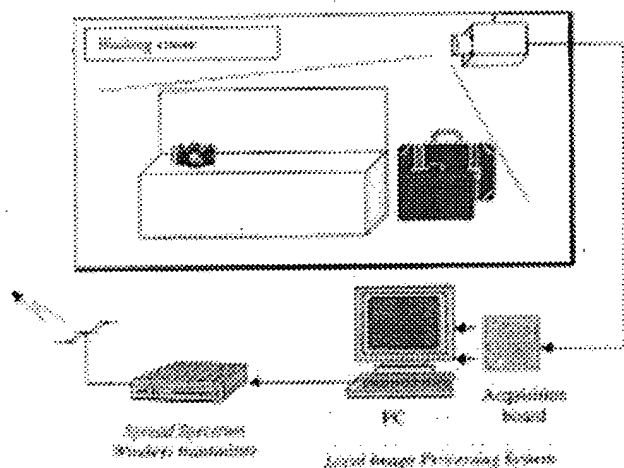


Fig. 2. Local processing system for abandoned object detection.

processing system (Section III), the Spread Spectrum-based multimedia communication system (Section IV), and the image processing at the remote control center (Section V) are fully described.

III. LOCAL PROCESSING SYSTEM FOR ABANDONED OBJECT DETECTION

A. Modular Structure of the Image Processing System

The architecture of a local image-processing system is shown in Fig. 3. The system is structured at different processing levels in order to provide a simple and flexible hierarchical architecture. A *module*, implementing a specific image-processing function, corresponds to each processing level [11]. The different

modules communicate to one another by exchanging processed multimedia information. From Fig. 3, it is easy to see that the information incoming from the sensors (i.e.; the monochromatic cameras) is processed step by step in order to produce the expected issue of the processing chain, i.e., the alert information to be sent to the remote control center. Each module has been implemented keeping into account the physical and/or virtual links existing between the current module and the previous and next ones in the considered chain.

B. Descriptions of the Single Modules

1) *Acquisition Module*: From the hardware point of view, the acquisition module consists of a low-cost acquisition board for real-time monochromatic and RGB-color image capture in PC-based video-surveillance applications. Exploiting the capabilities of the innovative *peripheral component interconnect* (PCI) bus architecture, the considered acquisition board can transfer the acquired data up to 45 Mb/s. The output of the acquisition module is a 256×256 -pixel digitized image, where each pixel is made up of eight information bits, corresponding to 256 gray levels.

2) *Change Detection Module*: This module is devoted to extracting “interesting” pixels from the acquired image sequences, when long-term changes are detected. Change detection can be considered as the most critical step for the proposed system. An efficient detection of the variations in the currently observed scene with respect to the background image is the basis for the development of real-time image-processing functions, implemented in the successive modules (see Fig. 3). The proposed algorithm [12] evaluates the difference between each pixel in the current frame and the corresponding pixel in the background image (i.e.; it operates at the *pixel level*), considering also the

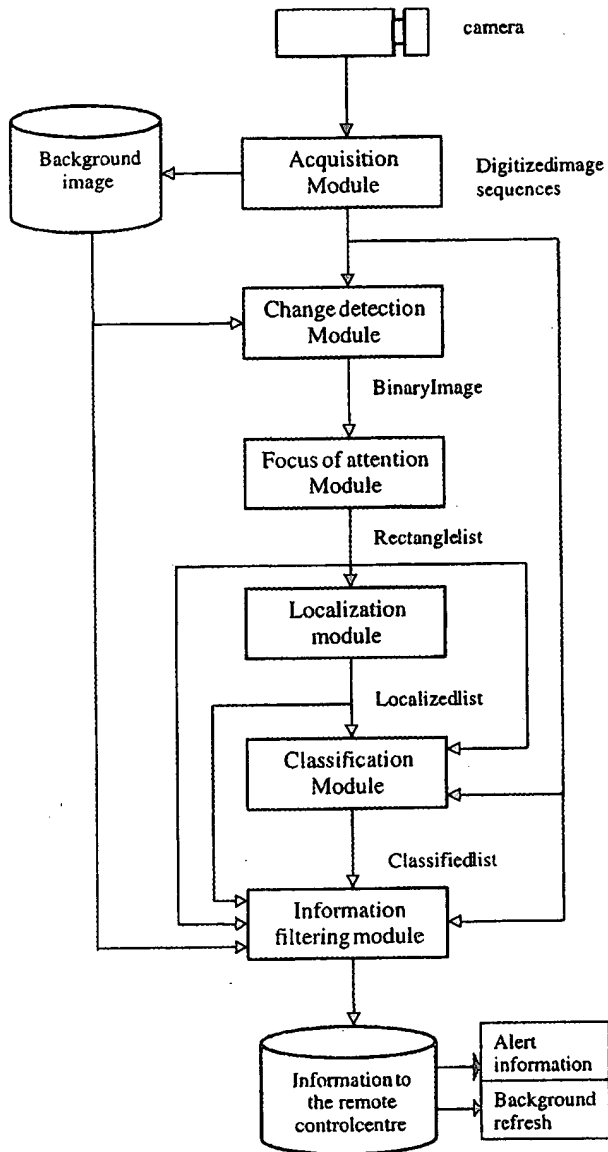


Fig. 3. Modular scheme of the local processing system for abandoned object detection.

permanence of the detected pixel's gray level in some consecutive frames: this makes the algorithm more robust against isolated noise peaks than the aforesaid simple difference. The implementation of the change detection module is based on the definition of an abandoned object, which is an object not belonging to the background scene and remaining in the same position for a long time [11]. Denoting by $p_k(x, y)$ the gray level of the pixel whose coordinates are (x, y) in the current frame k , and by $b(x, y)$ gray level of the corresponding pixel in the background image, the change detection algorithm considers the pixel (x, y) as a *possibly* changed pixel if the following condition holds:

$$|p_k(x, y) - b(x, y)| > S \quad (3.1)$$

where S is a threshold, defined on the basis of the lighting condition in the background image. The information concerning the simple difference between the corresponding pixels in the current frame and in the background image is not sufficient to deduce, in a reliable way, that a pixel has been changed by the presence in the scene of an object not belonging to the background. An impulsive noise peak may cause an incorrect change detection. For this reason, it is necessary to achieve also the luminance variations of the considered pixel in a fixed number of consecutive frames. Denoting by $p_{k+1}(x, y)$ the gray level of the pixel (x, y) , observed in the successive frame $k + 1$, the pixel can still be considered as a possibly changed pixel if the following condition holds:

$$|p_{k+1}(x, y) - p_k(x, y)| < S. \quad (3.2)$$

The condition expressed in (3.2) means that a pixel in the current image is to be regarded as a changed pixel, not only if its gray level is different from that of the corresponding pixel in the background image, but also if its gray level has a constant value in the successive frames. In order to perform an efficient change detection, for each pixel of the considered image two dynamic binary queues of length H are defined. H is the number of successive frame where the permanence of the gray-level value of the considered pixel is tested. The insertion/deletion of the queue items is managed by using a FIFO philosophy [12]. The two binary queues are defined as follows.

- $B_{x,y}$, which is the queue related to the difference between the pixel gray-level values observed in the current frame and in the background image, respectively (*frame-background difference*).
- $A_{x,y}$, which is the queue related to the difference between the pixel gray-level values observed in the successive frames (*frame-frame difference*).

The queue values are assigned by following the criteria defined as follows:

$$B_{x,y}(k) = \begin{cases} 1 & \text{if } |p_k(x, y) - b(x, y)| > S \\ 0 & \text{otherwise} \end{cases} \quad k = 1, \dots, H. \quad (3.3)$$

$$A_{x,y}(k) = \begin{cases} 1 & \text{if } |p_{k+1}(x, y) - p_k(x, y)| > S \\ 0 & \text{otherwise} \end{cases} \quad k = 1, \dots, H. \quad (3.4)$$

The decision on a possible pixel change is made only when the two queues are filled (i.e., after H acquired frames, which correspond to the time taken by the algorithm setup). In Fig. 4, an example of the queues inherent in change-detection algorithm considered is given.

The proposed algorithm considers the couples of values belonging to the two different queues, $B_{x,y}$ and $A_{x,y}$, as shown in Fig. 4. If the number of couples, whose binary values are $(1, 0)$, exceeds a fixed threshold t_a , the pixel (x, y) is regarded as a

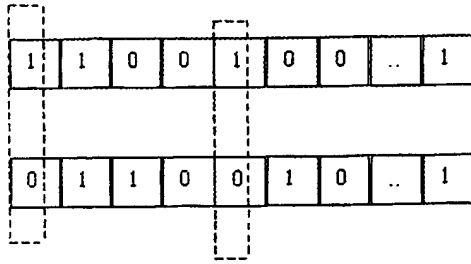


Fig. 4. Change detection module: queue management

changed pixel. The described module generates, as its output, a binary image C , where changed pixels are characterized by an assigned gray level equal to 255 (i.e., white color), whereas the unchanged pixels are characterized by a 0 gray level (i.e., black color). This is expressed by the formula:

$$C(x, y) = \begin{cases} 255 & \text{if } N_c(x, y) > t_a \\ 0 & \text{otherwise} \end{cases} \quad (3.5)$$

where $N_c(x, y) \triangleq \text{card}\{(B_{x,y}(k) = 1, A_{x,y}(k) = 0) : k = 1, \dots, H\}$ [12]. The optimal value of the threshold t_a is chosen in order to minimize the false-alarm probability P_{FA} , without increasing the misdetection probability P_{MA} . It is shown in [13] that considering a value $H = 16$ for the queue length results in an optimal value of $t_a = 7$. In Fig. 5(a)-(c), the background image, the current image and the binary image generated by the change detection module are presented, respectively.

3) Focus-of-Attention Module: This module aims at focusing attention on the areas of an image where meaningful changes have been actually detected. This allows the successive modules of the local processing system to consider only image areas where either a person or an object is really present in the guarded environment, thus making the computational load of the system less heavy. The input of the focus-of-attention algorithm is the binary image generated by the change detection algorithm. Residual noisy white pixels in the binary image are eliminated by using statistical morphological operators (i.e., statistical erosion and statistical dilatation) [13]. The output of the module is a list of *obstruction rectangles*, each characterized by the presence of a compact white area. To achieve this objective, the focus-of-attention algorithm first performs a *segmentation process* of the change areas detected by the change-detection module. This process allows one to separate two partially overlapped image regions. Then, the module generates a list of identified rectangular areas. The planar two-dimensional (2-D) coordinates (with respect to the background image) and the dimensions are provided for each obstruction rectangle.

4) Localization Module: This module aims at providing the coordinates of the obstruction rectangles in the three-dimensional (3-D) reference system related to the observed scene. The input of the module is the list of obstruction rectangles provided by the focus-of-attention module and the 3-D coordinates of the most significant planar regions in the scene, e.g., the floor, the walls, the tables, the benches, etc. The output of the module is

a list of *localized* obstruction rectangles; the spatial 3-D coordinates of each rectangle with respect to the guarded environment are provided to the successive modules. The algorithms implemented in the localization module are [14] the following:

- *camera calibration algorithms*, which aim at determining the relationship between the 3-D coordinates of a point in a spatial reference system and the planar 2-D coordinates of the same point in an acquired image. For the considered system, the calibration camera operations are performed off-line, as the camera is placed in a fixed position of the guarded room;
- *information transformation algorithms*. Once the camera calibration has been performed, it is possible by using 2-D and 3-D transformation algorithms to transform the information about the 2-D coordinates of a point in the image, so as to provide the coordinates of the same point in the 3-D reference system;

Thanks to these operations one can show the position of the detected region of interest in a map representing the guarded area.

5) Classification Module: The classification module is the most “intelligent” part of the system. In particular, it assigns each region of interest extracted by the previous modules (i.e., the localized obstruction rectangles) to one of the following four classes:

- *abandoned objects*, previously defined;
- *persons*: when a rectangle remains in the same position for some time, it is classified as a person;
- *lighting effects* (e.g., a localized variation in light due to an opened window);
- *structural changes* (e.g., a change in position of a chair).

The alarm is sent only when an abandoned object is recognized by the system. The use of a back-propagation neural network (BPN) [15] has been exploited in order to provide a reliable and real-time object classification.

6) Information Filtering Module: This module performs the *filtering* of the information to be transmitted to the remote control center. The implemented remote video-surveillance system has been developed in order to assist the human operator in monitoring some unattended railway stations. For this reason, the transmission of the information, which will be displayed on the monitors of the remote control center, is managed to ensure that the human operator will be really alerted whenever a potentially dangerous situation is detected by the system, without a decrease in the attention due to the transmission of a too large amount of video information. In order to fulfil with the bandwidth constraint on the communication network, the information sent to the remote control center must be exhaustive and not redundant. However, it must be sufficiently complete so that the human operator may easily realize the current situation at a remote guarded station. When the system is starting, only the fixed background image is transmitted to the remote control center. The permanence of the waiting-room background images on the monitors indicate to the human operator that: a) no abandoned object has been detected by the local processing system; b) no structural changes in the observed scene have occurred. When a structural changes occurs in the guarded scene (e.g., a change in

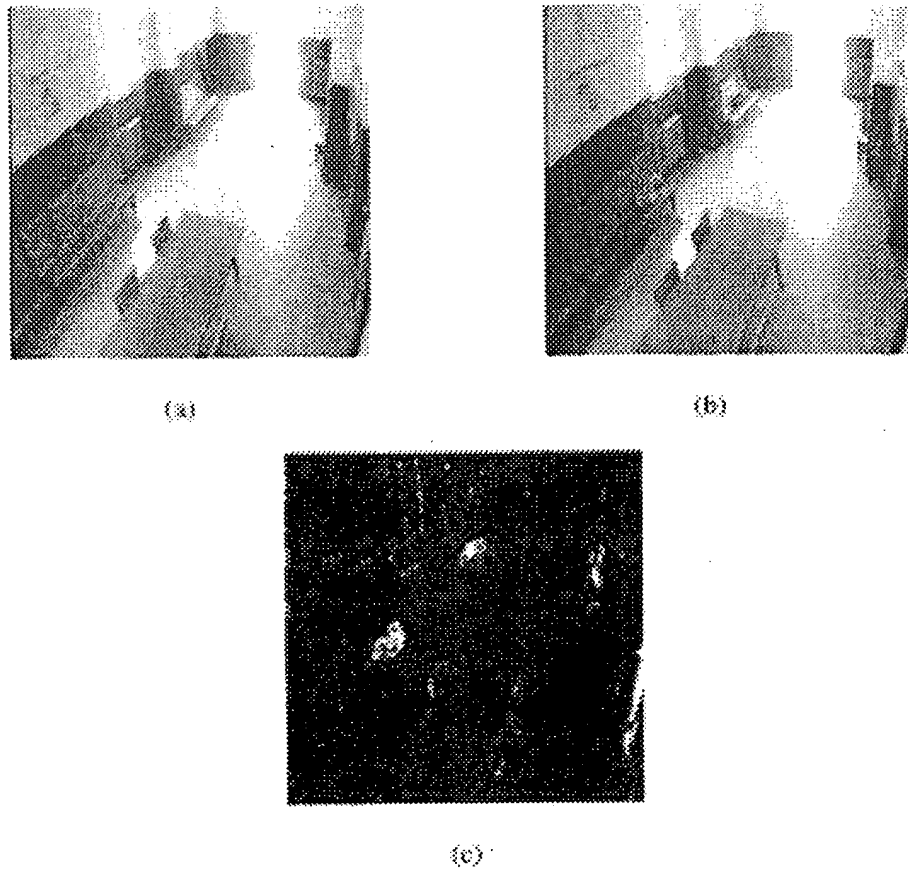


Fig. 5. Change detection module: (a) background input image, (b) current input frame, and (c) output binary image.

the position of a piece of furniture in the waiting room, etc.), the local processing system commands the retransmission of the updated background image (*background refresh*) [11]. The background image is a monochromatic $256 \times 256 \times 8$ bit image [see Fig. 5(a)]. The amount of information in the background image (before coding) is $D_{\text{back}} = 64$ KB. When a potentially dangerous situation is detected, the local processing system commands the transmission of the *alert information*, which consists of the following:

- the monochromatic image containing only a detected abandoned object. This small image is overlapped by the remote processing algorithms with the background image (see Fig. 6) [11] in order to form the complete scene concerning the alert situation. The average dimensions of rectangles containing the abandoned objects are equal to about 400 pixels. The amount of information about the abandoned-object image (before coding) is $D_{\text{alert}} = 3200$ bits. For the image-overlapping operation, the knowledge of the geometric position of the rectangle containing the abandoned object with respect to the background image is required. The amount of information about the 2-D coordinates of the center of the rectangle is $k_{2D} = 16$ bytes;
- the 3-D coordinates of the detected object. This information allows one to localize the abandoned object on the

map of the guarded environment. The size of such information is $k_{3D} = 16$ bytes.

IV. DS/CDMA MULTIMEDIA TRANSMISSION SYSTEM USING COMBINED SOURCE AND CHANNEL CODING TECHNIQUES

One of the most important problems concerning the transmission of multimedia information in remote AVS applications lies in uplink bandwidth constraints. As the most common networking applications (e.g., INTERNET, HDTV, Video on Demand, etc.) involve the transmission of a large amount of information from the network head-end to residential sites, really operating wireless and wired local area network (LAN) systems are generally characterized by an *asymmetrical* bandwidth availability, i.e., the bandwidth used for uplink (i.e., station-base) communications is generally much smaller than the bandwidth used for downlink (i.e., base station) ones. On the other hand, remote AVS applications need a high bandwidth availability over the uplink channel in order to ensure a real-time transmission of digitized image sequences from the guarded places to the remote control center [18]. The main problem concerning multimedia communications in AVS applications is the choice of an efficient source and channel coding strategy to ensure a real-time transmission of needed information, coping with reduced bandwidth resources over the uplink channels

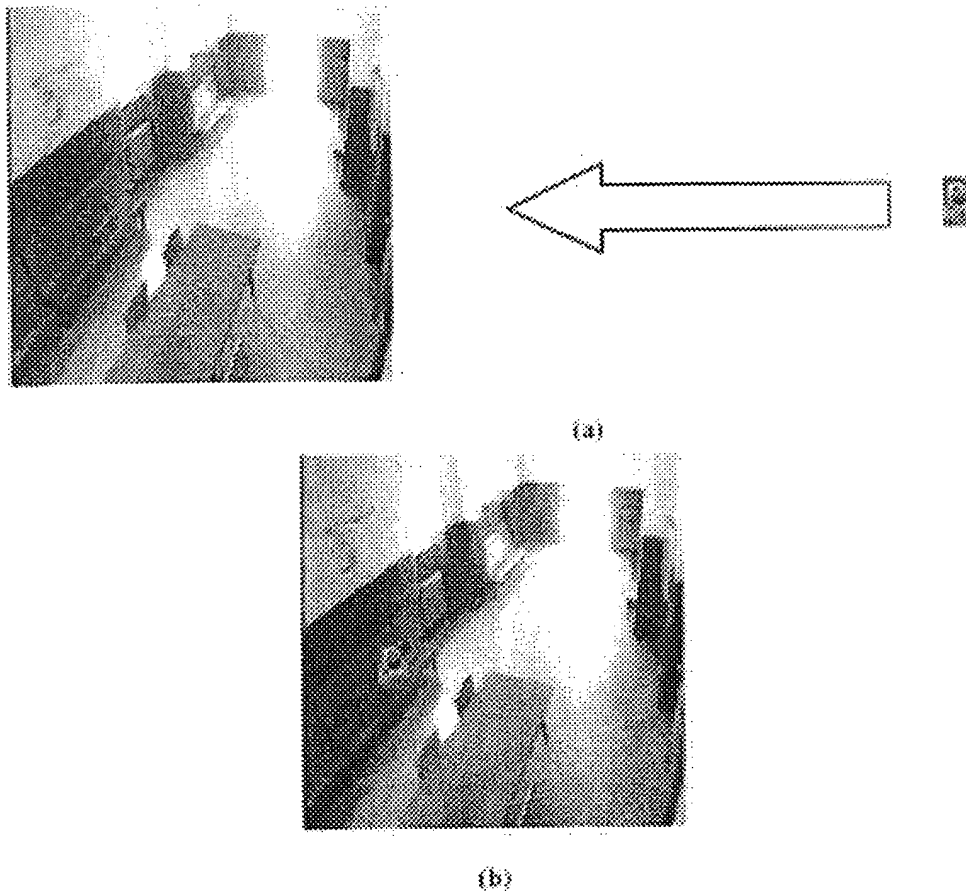


Fig. 6. Information filtering module: (a) filtered images transmitted and (b) received images overlap.

of really operating backbone networks. The communication system proposed in this work aims at providing a secure and noise-robust wireless transmission of the multimedia information related to the management of abandoned-object detection functionalities, keeping into account uplink bandwidth constraints. The described AVS communication system is based on the combination of advanced DS/CDMA channel-coding techniques with state-of-the art JPEG and forward error control (FEC) source coding techniques. The global architecture of the generic k th transmission $1 \leq k \leq K$ is shown in Fig. 7. Two 8-MHz DS/SS separate communication channels in the 2.4 GHz ISM [11] band are used for the background image transmission and for the alert information transmission respectively. All the K users can asynchronously send information over the two channels. The employment of the asynchronous DS/CDMA protocol seems the most suitable choice for the considered application, where the background-refresh operation and the alert information transmission involve a fully asynchronous access to the communication channel [11]. Moreover, the DS/CDMA wide-band transmission can ensure a more robust protection of transmitted data against the negative effects of noise, interference, multipath fading, accidental and/or intentional interception, and manipulation attempts [19].

The source coding of the above-described information involves two steps:

- Step 1) *Image Compression coding*. The images sent to the remote control center are compressed by using a JPEG standard encoder. The compression coding decreases the number of transmitted information bits, thus making it possible to increase the redundancy due error-correction coding and, especially, the processing gain of the DS/CDMA transmission system [19].
- Step 2) *Forward Error Control (FEC) coding*. It is known that even few and isolated bit errors can produce a very dramatic degradation of the received JPEG coded images, such as to impose the retransmission of the corrupted frames. A software algorithm for the detection and correction of transmission errors occurring in the JPEG-coded bit streams is presented in [16]. The proposed method is quite efficient and allows very high-quality images to be recovered from the corresponding corrupted JPEG ones. However, this algorithm can increase the computational complexity of the whole system, thus compromising real-time processing requirements. In our work, a forward error concealment strategy [23] is applied by using a FEC coding. In order to improve the performance in terms of reduced bit error rate (BER), without increasing information re-

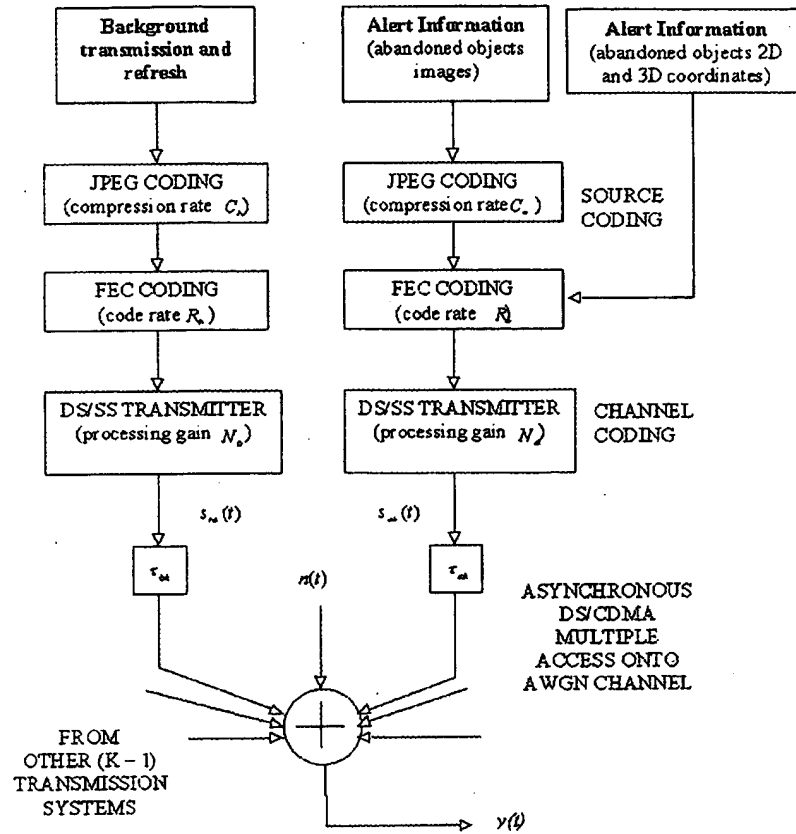


Fig. 7. Global scheme of the multimedia transmission system using combined source and channel coding techniques for the radio-link communication between unmanned station and remote control center.

dundancy too much, low-rate *convolutional Viterbi coding* [17] has been employed for the proposed system.

The block diagram of the DS/CDMA communication system is presented in Fig. 8. The same diagram is adopted for both the background channel and the alert-information channel. The only difference between the two modem schemes consists in the DS/SS processing-gain N , which is the length of the pseudonoise (PN) codes employed to spread the spectrum of the transmitted signal.

The processing-gain is the key parameter for the project of an asynchronous DS/CDMA transmission system. It is known that N is the *power gain* (generally expressed in dB) achieved by the received signal with respect to all kinds of interfering signals [19], which are:

- narrowband and wide-band signals transmitted over the same 8-MHz channel by other users, but not belonging to the same DS/CDMA communication system used;
 - narrowband impulsive noise due to environmental electromagnetic emissions (ingress noise) and/or attempts at intentional interference (e.g., jamming pulses [19]);
 - wide-band multiple-access interference (MAI) due to the nonidentical orthogonality of the PN spreading sequences employed by each user of the DS/CDMA system [20].
- The presence of MAI, which is generally a non-Gaussian

interference, involves some well-known problems in the DS/CDMA system design, in terms both of a reduction in the global capacity of the system, and a correct evaluation of BER performances when the Gaussian approximation for the MAI pdf is not acceptable (e.g., in the case of few users [21]). The choice of a suitable value of the processing gain can reduce the effects of MAI, thus allowing a considerable number of users to access the same bandwidth with a negligible performance degradation in terms of BER.

The processing-gain dimensioning for the two DS/CDMA channels used (i.e., background channel and alert-information channel) should take into account the tradeoffs between bandwidth constraints, and requirements in terms of quality of service (QoS). The mathematical expressions for the processing-gain for the background channel N_{back} and for the processing-gain for the alert information channel N_{alert} are:

$$N_{\text{back}} = \frac{B_{av} C_{\text{back}} R_{\text{back}} t_{\text{back}}}{D_{\text{back}}} \quad (4.1)$$

$$N_{\text{alert}} = \frac{B_{av} R_{\text{alert}} t_{\text{alert}}}{\left(\frac{D_{\text{alert}}}{C_{\text{alert}}} + k_{2D} + k_{3D} \right)} \quad (4.2)$$

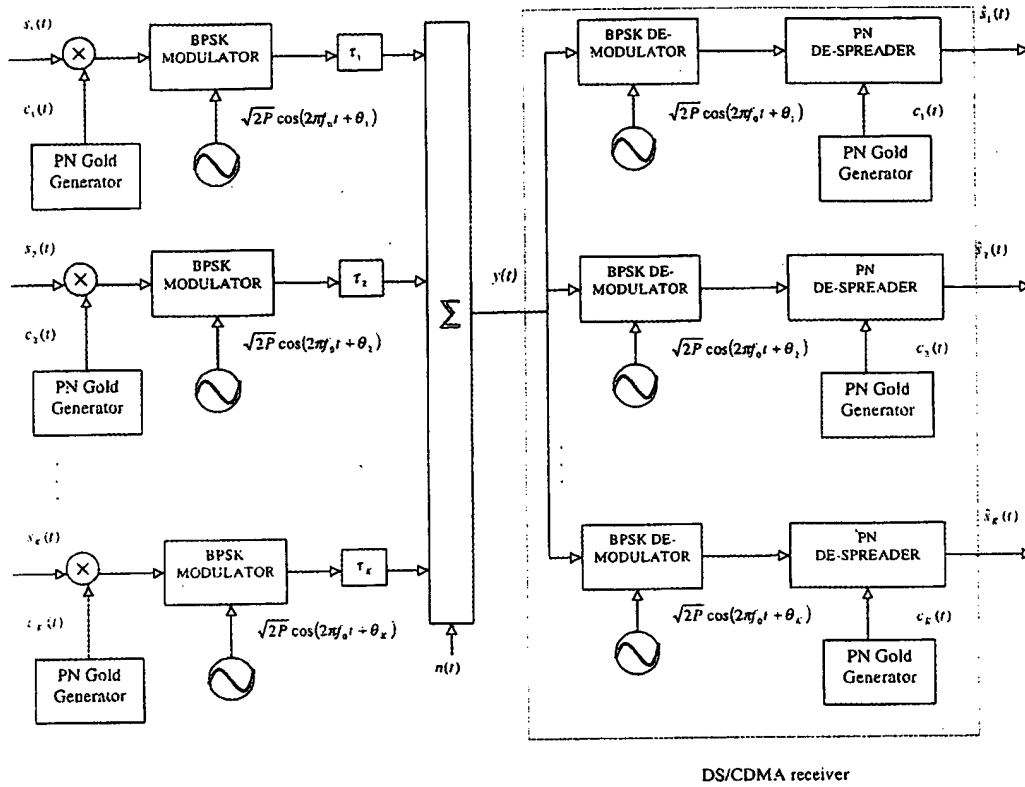


Fig. 8. Block diagram of the DS/CDMA digital transmission link.

where

- B_{av} bandwidth available for the transmission over the two channels used (i.e., 8 MHz);
- C_{back} and C_{alert} JPEG compression coefficients for the two different kinds of multimedia information transmitted (i.e., background image and alert image);
- R_{back} and R_{alert} FEC code rates for the background transmission and for the alert-information transmission, respectively;
- t_{back} and t_{alert} times taken for transmitting the background image and the alert-information image, respectively.

In order to meet the real-time functional specifications of the system, the following values of the transmission times have been considered: $t_{back} = 1$ s, and $t_{alert} = 500$ ms.

Two different FEC codes at different rates $R_{back} = 1/2$ and $R_{alert} = 1/3$ have been chosen for the transmission of the background image of the alert-information image, respectively. The choice of different code rates for the different information, to be transmitted to the remote control center has been made in order to protect, in a more robust way, alert information, which is more critical from a safety point of view (i.e., a retransmission of the background image is surely more acceptable a retransmission of the image of an abandoned object). The JPEG code rates have been chosen in order to reduce the transmission bit rate, without a significant degradation of the quality of the related decoded images. For this reason, a compression rate $C_{back} = 16$ has

been adopted for the background image, and a compression rate $C_{alert} = 10$ has been adopted for image of an abandoned object. At a glance, the choice of using the JPEG coding also for the alert-image compression might seem quite strange, as the average dimensions D_{alert} of the image of an abandoned object is very small, as compared with the ones of the background image. The basic reason for this choice is the necessity for providing a very robust FEC coding and a high value of the processing-gain for critical alert information in order to decrease the system BER even in the presence of very noisy channels. The low-rate JPEG coding chosen should not compromise the quality of the decoded image of an abandoned object. For the considered numerical values of the source-coding rates and the for the defined time and bandwidth constraints, the processing-gain values achieved for the background channel and for the alert information channel are $N_{back} = 127$ and $N_{alert} = 2047$, which are suitable values to support a considerable number of users transmitting over the two channels, without a significant degradation of the BER performances even for low signal-to-noise ratio (SNR) values, as shown in Section VI.

V. IMAGE PROCESSING AT THE REMOTE CONTROL CENTER

The remote control center of the system is placed in an urban railway station and can be managed by a human operator. The DS/CDMA matched filter receiver in Fig. 1, (the related block diagram is shown in Fig. 8), receives the RF DS/CDMA signals from the channel, and performs the BPSK demodulation and despreading of source-coded multimedia signals. A PC-based

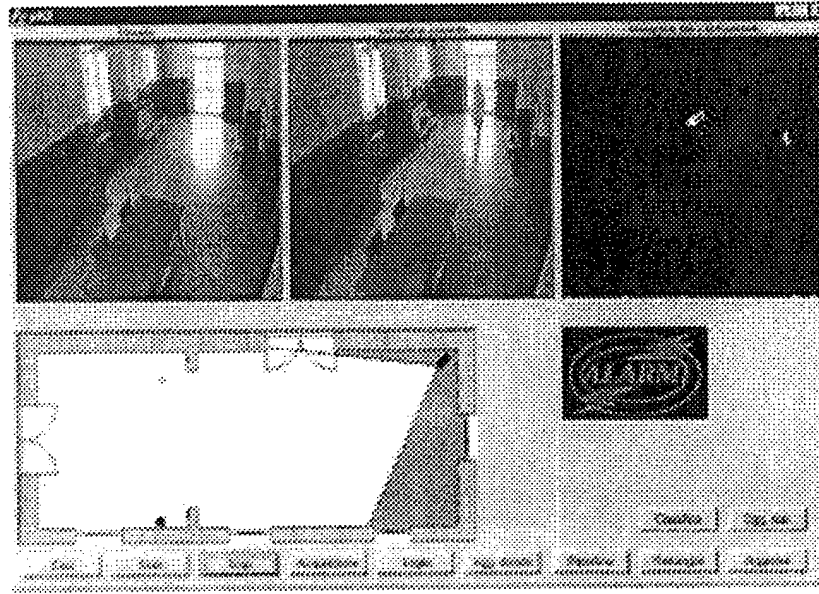


Fig. 9. Man-machine interface at the remote control center.

TABLE I
PERFORMANCES IN TERMS OF CORRECT DETECTION, FALSE ALARM, AND MISDETECTION PROBABILITIES PROVIDED BY THE LOCAL IMAGE PROCESSING SYSTEM

	Abandoned object	Person	Lighting effects	Structural Change
P_{DET}	99%	86.4%	82.6%	99%
P_{MA}	1%	13.6%	17.4%	1%
P_{FA}	2.6%	2.2%	4.8%	2%

central processing architecture, named "server" (see Fig. 1), is devoted at performing the following operations.

- *Source decoding of the multimedia bitstreams received*, i.e., FEC and JPEG decoding of both the background image and the alert-information transmitted by each user. When a JPEG decoding operation fails due a fatal error, the source-decoding module issues an automatic retransmission signal to the local processing system. The retransmission can be issued by means of easy visual commands, also when the quality of the visualized images is strongly degraded;
- *Presentation of alarm situations to the human operator*. For each guarded station, the data concerning the implemented video-surveillance functionalities are displayed on a particular monitor through a suitable man/machine interface. The man/machine interface used for the local monitoring of the railway station of Borzoli, near Genoa (Italy) is shown in Fig. 9. In the upper left corner of the interface, the background image of the guarded waiting-room is presented to the operator. In the case of an alarm situation, the received images containing the abandoned objects detected by the local processing system are overlapped with the background image and shown to the operator on the upper middle side of the interface, together with a clearly visible alarm signaling. The abandoned objects are also positioned on a 2-D map placed in the lower left side of the interface.

The HW/SW remote processing architecture can be effectively implemented by using a PC-based high performance computing network (HPCN) architecture. The HPCN is so structured: one PC is the control station for the operator and the other ones are devoted to the remote processing of the information transmitted by each guarded station. The connection among the PCs is provided by a FastETHERNET network (transfer rate up to 100 Mb/s). The use of the WINDOWS NT 4.0 operating system can ensure a reliable management of the whole architecture.

VI. NUMERICAL RESULTS

A demonstrator of the local image-processing system for abandoned-object detection, (the software modules of the system have been described in Section III), has already been realized in our laboratories and subsequently tested by using image sequences acquired from the waiting-room of the aforesaid railway station. The achieved performances of the system in terms of correct detection (P_{DET}) misdetection (P_{MA}) and false alarm (P_{FA}) probabilities for each class of detected objects and/or changes, listed in Section III are shown in Table I. These results were obtained by comparing the output of the classification module with a direct observation of the scene considered.

It is worth noting that the best performances achieved in terms of P_{DET} are related to the class of abandoned objects; the only

TABLE II
PERFORMANCES IN TERMS OF BER VERSUS SNR PROVIDED BY THE SIMULATED DS/CDMA BACKGROUND IMAGE TRANSMISSION SYSTEM

	Abandoned object	Person	Lighting effects	Structural Change
P_{DET}	99%	86.4%	82.6%	99%
P_{FA}	1%	13.6%	17.4%	1%
P_{MISS}	2.6%	2.2%	4.8%	2%

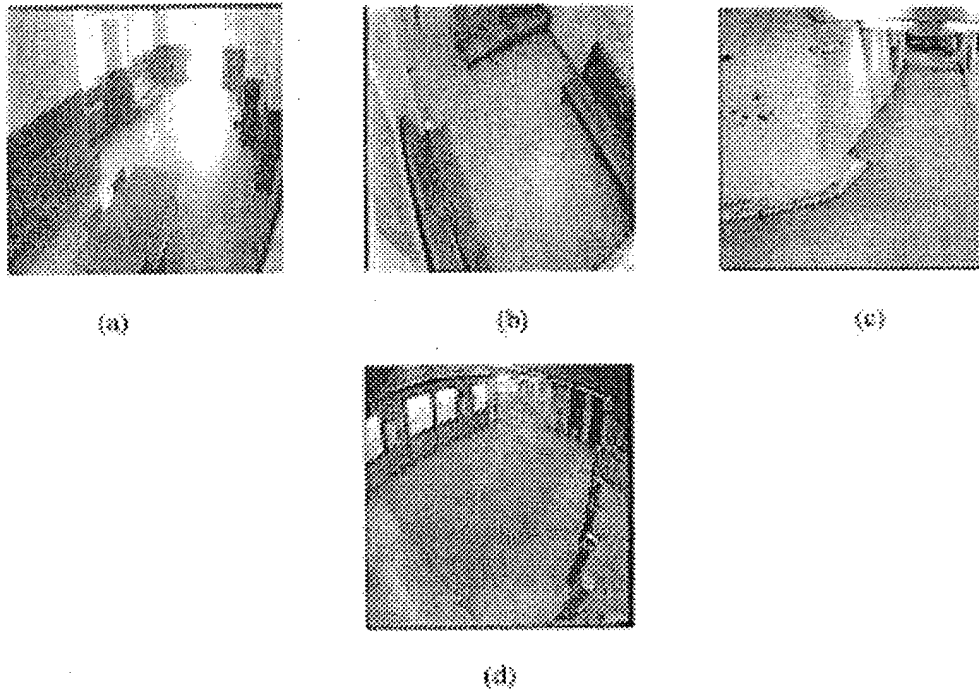


Fig. 10. Simulation of DS/CDMA background transmission: (a) background image transmitted by reference user; (b), (c), (d) background image transmitted by the interfering users.

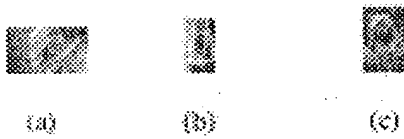


Fig. 11. Simulation of DS/CDMA abandoned object image transmission: (a) abandoned object image transmitted by the reference user; (b), (c), abandoned object image transmitted by the interfering users.

detection of an abandoned object requires that an alarm situation is notified to the human operator. The data shown in Table I were obtained after the camera calibration and the neural-network training performed for the test environment. Concerning the performances of the system in terms of processing time, the total taken time for the execution of the whole processing chain shown in Fig. 3 was $t_{EL} = 1.15$ s. This performance has been obtained by using a PC-based hardware/software architecture, with CPU PENTIUM INTEL 200 MHz, 64 Mb of RAM capacity, PCI bus, and a WINDOWS 95 operating system. Such performances meet the real-time processing requirements of the system. The multimedia DS/CDMA transmission system described in Section IV has been studied for feasibility and sim-

ulation aspects. The implementation of an innovative simulator of the entire DS/CDMA transmission system depicted in Fig. 8 has been presented in [22]; the SIMULINKTM libraries have been exploited, working in the MATLAB 5.2 environments. A four-user DS/CDMA transmission system has been considered, with Gold spreading sequences of length $N = N_{back} = 127$ for the background channel and $N = N_{alert} = 2047$ for the alert-information channel. The simulations of a four-user DS/CDMA transmission of background images and the simulation of a three user transmission of abandoned-object images were performed. The background images of the waiting rooms of four unattended railway stations transmitted by the four users of the DS/CDMA system are shown in Fig. 10(a)–(d). The abandoned-object images transmitted by three users are presented in Fig. 11(a)–(c). The images contained in Figs. 10(a) and 11(a) are considered as the transmitted reference information to be despread and decoded, whereas the other images are regarded as interfering bit-streams. All the simulations were performed assuming the hypothesis of an AWGN channel.

The BER performances achieved by the simulations of the background-image transmissions are shown in Table II, together with the number of noise-altered JPEG coefficient.

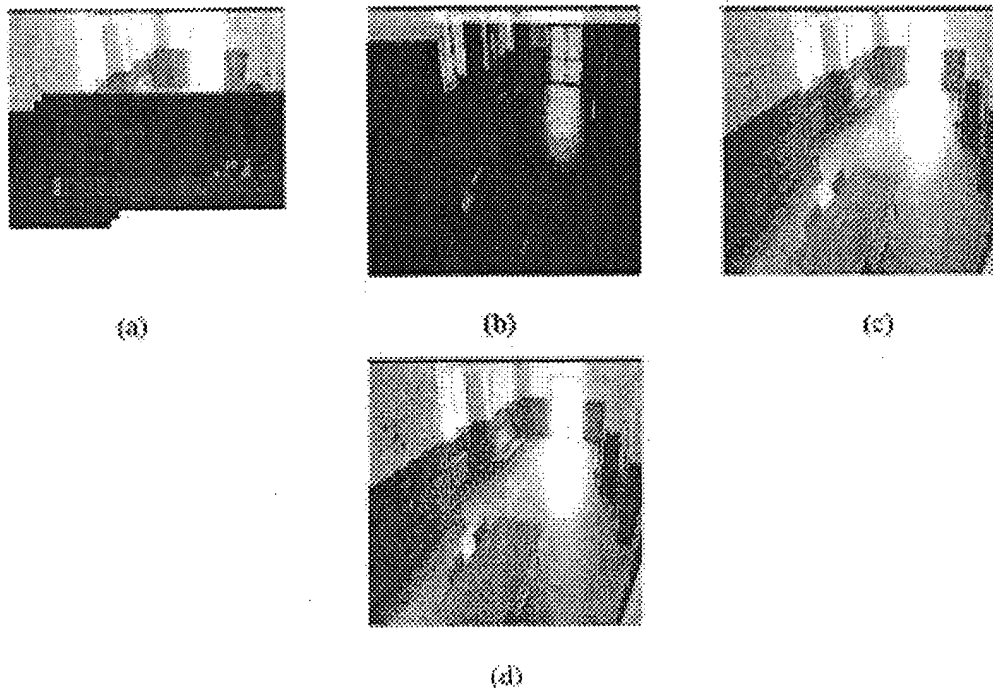


Fig. 12. Results of DS/CDMA transmission simulation (received and decoded reference user's background image): (a) transmission without FEC coding and SNR = 9 dB, (b) transmission without FEC coding and SNR = 10 dB (first simulation), (c) transmission without FEC coding and SNR = 10 dB (second simulation), and (d) transmission with convolutional FEC coding and SNR = 6 dB.

Just to demonstrate the importance of the introduction of the source FEC coding into the considered system, from Table II one can see that an error-free detection of the JPEG coded bitstream was obtained at a very low SNR (i.e., SNR = 6 dB), whereas the results of the simulations performed without the FEC coding show that a considerable number of noise-altered JPEG coefficients is resulting from higher SNR values (e.g., 8 and 9 dB). Two simulations for SNR = 10 dB, and without the FEC coding were performed. The BER achieved and the number of altered JPEG coefficients were the same for the two simulations, but the qualities of the decoded images were very different (see Fig. 12).

There is not any decoded image for what concerns the simulation without FEC coding and with SNR = 8 dB. This is due to some fatal errors on the received JPEG bitstream, which avoided a direct decoding. Fig. 12(a) shows the decoded image obtained by the simulation without FEC for SNR = 9 dB. In this case, the JPEG bitstream was decoded, but the resulting background frame was useless. This happened also for the first simulation without FEC and for SNR = 10 dB [see Fig. 12(b)]. The only noise-altered JPEG coefficient was placed in a critical position in the received bitstream, thus causing the failure of the decoding process. On the contrary, the second simulation without FEC and for SNR = 10 dB provided a very good quality of the decoded image, shown in Fig. 12(c), as the altered JPEG coefficient was not placed in a critical position in the received bitstream (it is the last JPEG coefficient). Even a single bit error may cause the failure of the JPEG decoding process, thus requiring the retransmission of a frame. An error-free four-user DS/CDMA transmission over

the background channel is reached for SNR = 11 dB, without using any FEC coding. The same performance was achieved for SNR = 6 dB by using the 1/2-rate convolutional FEC coding, as described in Section IV. The decoded image resulting from this simulation is presented in Fig. 12(d).

Concerning the DS/CDMA transmission of abandoned-object images, an error-free bitstream was obtained for SNR = 7 dB, without using any FEC coding. This result proves the expected capability of long spreading codes to provide a strong reduction of the multi-access interference also at low SNR values. The use of the 1/3-rate convolutional FEC code can surely improve the robust protection already provided by the DS/SS channel coding against all kinds of noise.

A. Color Image Processing for Abandoned Object Detection

The change detection module of the local image processing system has been originally designed to process images acquired from a monochromatic camera. Full exploitation of color information in the change detection algorithm could imply some considerations related both to image processing and to communication aspects.

From the image processing point of view, some tests performed about the use of color image sequences revealed that the processing time per frame achieved by employing a change detection algorithm quite similar to the approach shown in [24] increases up to 2.0 s (i.e., about 1.5 times greater than the time achieved using monochromatic images), whereas the false alarm probability concerning the abandoned object detection becomes

equal to 2% (instead of 2.6%, as shown in Table I) and the related misdetection probability becomes equal to 0.6% (instead of 1%).

For what concerns the design of the communication system, as the source bit rate would triplicate with respect to the monochromatic case in both the transmission channels, the bandwidth constraints expressed in Section IV could not stand more without modifying the parameters of source and channel coding (i.e., JPEG compression rates, FEC code-rates and spreading factors). Such modifications would be finalized to allow a faster transmission than the one foreseen for the monochromatic images over the same 8-MHz bandwidth, thus involving weaker coding and minor protection against channel impairments. This fact could imply a significant degradation of the quality of the results displaying to the human operator working inside the remote control center. Otherwise, the coding parameters of the communication system could be unmodified, if a bandwidth equal to three times the 8-MHz bandwidth foreseen (i.e., 24 MHz) would be available in the uplink direction. However, as mentioned in Section I, the uplink bandwidth resources are generally quite scarce. One can see that the use of color image sequences can improve the robustness of the local image processing system, already tested with satisfactory results when monochromatic images are employed (see Table I). On the other hand, the costs to be paid for such improvement concern with a slight increase of the processing time (however acceptable for the considered application) and a consistent increase of the bandwidth to be occupied for the transmission of the multimedia information to the remote control center. Otherwise, if the bandwidth constraints were respected, a decrease of QoS related to the results collecting and presentation at the remote control center could be the tradeoff to be accepted.

VII. CONCLUSION

In this paper, a distributed video-surveillance system for monitoring unattended railway stations has been presented. The "intelligence" of the system has been distributed by implementing a real-time local processing architecture, which is devoted to acquiring images from the guarded environments and to processing such images to filter areas containing abandoned objects. The efficiency of a prototype for the local processing system in terms of low false-alarm and misdetection probabilities has been proved by tests performed in both our laboratory and real railway environments.

The communication system, based on state-of-the-art combined source and channel coding techniques and on asynchronous DS/CDMA multiple access techniques, allows some users (i.e., unattended railway stations) sharing the same 8-MHz bandwidth portion to perform a secure and noise-robust transmissions of background images and alert information to a remote control center. The image processing at the remote control center is reduced to an efficient decoding and collection of the multimedia information transmitted by each user of the system and to the presentation of this information to a human operator. These suitable characteristics allow the exploitation of such a system for the remote monitoring of a wide range

of unattended environments (e.g., supermarkets, airports, car parks), thus, not limiting its use to railway environments.

ACKNOWLEDGMENT

The authors wish to thank E. Stringa for her valuable contribution to the paper.

REFERENCES

- [1] P. G. Michalopoulos, "Vehicle detection video through image processing: The AUTOSCOPE system," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 21–29, Feb. 1991.
- [2] J. M. Blossville, J. M. Morin, S. Bouzar, F. Lenoir, and P. Lochegnie, "Results of the TRISTAR AID system obtained from 3 field experiments," in *Proc. 7th Int. Conf. Road Traffic Monitoring and Control*, London, U.K., April 1994, pp. 168–171.
- [3] C. E. Smith, C. A. Richards, S. A. Brand, and N. P. Papanikolopoulos, "Visual tracking for intelligent vehicle-highway systems," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 744–759, Nov. 1996.
- [4] E. J. Charlson, H. L. Graham, and X. Zhuang, "Video camera method for moving-vehicle location," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 686–693, May 1998.
- [5] N. Kehtavarnaz, N. Griswold, K. Miller, and P. Lescoe, "A transportable neural-network approach to autonomous vehicle following," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 694–702, May 1998.
- [6] G. L. Foresti, V. Murino, and C. S. Regazzoni, "Vehicle recognition and tracking from road image sequences," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 301–318, Jan. 1999.
- [7] J. P. Deparis, S. A. Velastin, and A. C. Davies, "CROMATICA project: A collection of telematic tools for improvement of safety in public transport," in *Advanced Video-Based Surveillance Systems*, C. S. Regazzoni, G. Fabri, and G. Vernazza, Eds. Norwell, MA: Kluwer, 1999, ch. 5, pp. 201–212.
- [8] A. Teschioni and C. S. Regazzoni, "Performance evaluation strategies of an image processing system for surveillance applications," in *Advanced Video-Based Surveillance Systems*, C. S. Regazzoni, G. Fabri, and G. Vernazza, Eds. Norwell, MA: Kluwer, 1999, ch. 2, pp. 76–90.
- [9] G. L. Foresti, C. S. Regazzoni, and A. Teschioni, "Progressive image coding for visual surveillance applications based on statistical morphological skeleton," *Proc. European Signal Processing Conf. (EUSIPCO)*, vol. II, pp. 1159–1161, Sept. 1996.
- [10] G. L. Foresti and C. S. Regazzoni, "A new shape representation method for progressive image coding at a very low bit rate," *IEEE Signal Process. Lett.*, vol. 4, pp. 267–281, Sept. 1997.
- [11] E. Stringa, C. Sacchi, and C. S. Regazzoni, "A multimedia system for the surveillance of unattended railway stations," in *Proc. European Signal Processing Conf. (EUSIPCO)* Rhodes, Greece, Sept. 1998, vol. III, pp. 1709–1712.
- [12] C. S. Regazzoni, A. Teschioni, and E. Stringa, "A long term change detection method for surveillance applications," in *Proc. 9th Int. Conf. ICIAF* Florence, Italy, Sept. 1997, vol. I, pp. 485–492.
- [13] G. L. Foresti, C. Regazzoni, A. N. Venetsanopoulos, and G. Vernazza, "Statistical morphological filters for binary image processing," in *Proc. IEEE International Conf. Acoustic, Speech and Signal Processing (ICASSP)* Adelaide, U.K., April 1994, vol. 5, pp. 77–80.
- [14] G. L. Foresti, V. Murino, and C. S. Regazzoni, "Real time adaptive regulation of a visual camera for automatic investigation of changing environments," in *Proc. 19th IEEE Int. Conf. Industrial Electronics—IECON*, Hawaii, Nov. 1993, pp. 1633–1638.
- [15] E. Stringa, A. Teschioni, G. Vernazza, and C. Regazzoni, "PFT2 research project: Multisensory image and signal interpretation for the surveillance of unattended railway stations—Classification by using neural networks," (in Italian), Univ. of Cagliari, Dept. Electrical and Electronic Engineering, Tech. Rep. DIEEE/PFT2/96.001 30.PF74/01.97/VM, Jan. 1997.
- [16] Y. Han and J. Leou, "Detection and correction of transmission errors in JPEG images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 221–231, April 1998.
- [17] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995, ch. 8, pp. 413–528.
- [18] P. Mahonen, "Integration of wireless networks and AVS," in *Advanced Video Based Surveillance System*, C. S. Regazzoni, G. Fabri, and G. Vernazza, Eds. Norwell, MA: Kluwer, 1999, ch. 4, pp. 144–153.

- [19] R. L. Pickholtz, D. L. Schilling, and L. B. Milstein, "Theory of spread-spectrum communication—A tutorial," *IEEE Trans. Commun.*, vol. COM-30, pp. 856–884, May 1982.
- [20] M. B. Pursley, "Performance evaluation for phase-coded spread spectrum multiple-access communication—Part I: System analysis," *IEEE Trans. Commun.*, vol. COM-25, pp. 795–799, Aug. 1977.
- [21] A. Teschioni, C. Sacchi, and C. S. Regazzoni, "Non-Gaussian characterization of DS/CDMA noise in few-user systems with complex signature sequences," *IEEE Trans. Signal Processing*, vol. 47, pp. 234–237, Jan. 1999.
- [22] P. Piccardo, C. S. Regazzoni, C. Sacchi, G. Sciani, and A. Teschioni, "Software design and simulation of a DS/CDMA multimedia transmission system for remote video-surveillance applications," in *Multimedia Communications*, F. De Natale and S. Pupolin, Eds. New York: Springer-Verlag, 1999, ch. 5, pp. 557–568.
- [23] Y. Wang and Q. F. Zhu, "Error and concealment for video communication: A review," *Proc. IEEE*, vol. 86, pp. 974–997, May 1998.
- [24] A. Teschioni and C. S. Regazzoni, "A robust method for reflection analysis in color image sequences," in *Proc. European Signal Processing Conf. (EUSIPCO) 1998* Rhodes, Greece, Sept. 1998, vol. IV, pp. 2165–2168.



Claudio Sacchi was born in Genoa, Italy, in 1965. He received the Laurea degree in electronic engineering and information science from the University of Genoa in January 1992.

In January 1996, he joined the Signal Processing and Telecommunications Group of DIBE, where he worked on the application of spread-spectrum techniques to railway signaling systems. During 1997, he carried on his research activity under grant from Italian National Research Council (CNR), within the framework of the PFT2 Transport

Programme, focusing his work on the design and implementation of remote video-based surveillance systems for railway transport applications. From 1998 to 1999, he worked in the technical management of the EU-ESPRIT AVS-RIO project related to the implementation of a cable network-based video-surveillance of tourist sites. Currently, he is working with CNIT (National Italian Inter-University Consortium for Telecommunications) in national project activities, related to the application of CDMA techniques to multimedia satellite communications. His main research interests concern with signal and image processing techniques for multimedia communications over wireless and wired networks and remote video-based surveillance applications.

Dr. Sacchi is a member of AEI.



Carlo S. Regazzoni (M'89–SM'00) was born in Savona, Italy, in 1963. He received the Laurea degree in electronic engineering and the Ph.D. degree in telecommunications and signal processing from the University of Genoa, Genoa, Italy, in 1987 and 1992, respectively.

In 1987, he joined the Department of Biophysical and Electronic Engineering (DIBE), University of Genova, where he is responsible for the Signal Processing and Telecommunications Research Group (SP&T). He has been an Assistant Professor in Telecommunications at DIBE since 1995. Since 1998, he has been Professor of Telecommunication Systems in the Engineering Faculty of the University of Genova. Since 1997 and 1998, he has been a Visiting Professor on pattern recognition at master course in information technology at Cogefo, Milan, Italy, and a Visiting Professor at the University of Trento, Italy, respectively. He has been responsible in several EU research and development projects and in several research contracts with Italian industries. His current main research interests are multimedia and nonlinear signal and video processing, signal processing for telecommunications, multimedia broadband access techniques, and computer vision for video-based surveillance. He served as a reviewer for international journals and for EU in different research programs. He is coeditor of the book *Advanced Video-Based Surveillance Systems* (Norwell, MA: Kluwer, 1999). He is author or coauthor of 38 papers in international scientific journals and of more than 120 papers at refereed international conferences.

Dr. Regazzoni has been Chairman of an International Workshop on Advanced Video Based Surveillance, Genova in 1998 of Special Sessions in the same field at the International Conference on Signal Processing (Eusipco2000), Tampere, Finland, 2000. He is a member of IAPR.

Scanning the Issue/Technology

Special Issue on Video Communications, Processing, and Understanding for Third Generation Surveillance Systems

I. INTRODUCTION

A surveillance system can be defined as a technological tool that assists humans by providing an extended perception and reasoning capability about situations of interest that occur in the monitored environments. Human perception as well as reasoning are constrained by the capabilities and limits of human senses and mind to simultaneously collect, process and store limited amount of data. For example:

- only information coming from a limited spatial area can be directly sensed and processed by the human at a given time;
- the complexity of the situations that can be analyzed is usually limited to events, occurring at different time instants, that can be associated by reasoning with their common causes.

Surveillance systems provided varied degrees of assistance to humans evolved in an incremental way according to the progress in surveillance technologies [1]. We will describe in the following sections the details of the successive generations of surveillance systems that increasingly utilize a larger set of sensors as well as more flexible and robust processing strategies.

This Special Issue focuses on the problems of last generation surveillance systems and highlights solutions to these problems that are based on a stronger integration of techniques for multisensor data acquisition, communications and processing. This integration is possible by the common "full digital" perspective on which the techniques used by new systems are based. Next generation surveillance systems can be considered as an emerging application field requiring multidisciplinary expertise going from signal and image processing, to communications and computer vision. This multidisciplinary view is common to many applications in the information and communications technology (ICT) domain, such as videoconferencing, ambient intelligence, etc. There is a growing interest in surveillance applications due to the growing availability of cheap sensors and processors at reasonable costs. There is also a growing need from the public for improved safety and security in large urban environments

and improved usage of resources of public infrastructure. This, in conjunction with the increasing maturity of algorithms and techniques, is making possible the application of this technology in various application sectors such as security, transportation, and the automotive industry. In particular, the problem of remote surveillance of unattended environments has received growing attention in the last years, especially in the context of:

- a) safety in transport applications [2], [3], such as monitoring of railway stations [4], [5], underground stations [6], [7], airports [8]–[10] and airplane routes [11]–[13], motorways [14], [15], urban and city roads [16]–[23], maritime environments [24]–[27];
- b) safety or quality control in industrial applications, such as monitoring of nuclear plants [28] or industrial processing cycles [1]–[3];
- c) improved Security for people lives, such as monitoring of indoor or outdoor environments like banks [29], supermarkets [6], car parking areas [30], waiting rooms [31], buildings [32], [33], etc., remote monitoring of the status of a patient [34], remote surveillance of the human activity [35]–[47];
- d) military applications for surveillance of strategic infrastructures [48], [49], enemy movements in the battlefield [50], [51], air surveillance [52], [53].

In order to satisfy a market potentially so large, strong research innovations are required that allow surveillance engineers and end-users to take advantage of innovative communication solutions, processing, and understanding methods that are developed by researchers. The goal of this Special Issue is to point out the key aspects and technological trends of the last generation of surveillance systems.

While several modalities of sensing such as audio, video, and chemical sensors are useful in monitoring; we chose to concentrate on those applications where visual information plays the most important role. Video communication, processing, and understanding can be considered as a fundamental modality for surveillance applications.

This is due to several factors.

- Temporally organized visual information is the major human source of information about the surrounding environment.

- As the number of cameras increase, event monitoring by personnel is rather boring, tedious, and error-prone. The automatic preprocessing of the video information by a surveillance system can act as a prefilter to human validation of the events. Thus, it is a natural mechanism to manage the complexity of monitoring a large site. In addition, a high-level interface presenting the events in a site is a most user-friendly and widely acceptable presentation.
- The cost of the video sensor is considerably lower compared to other sensors when one takes into account the area of coverage and event analysis functionality provided by using video as the sensing modality for monitoring.
- A large body of knowledge exists in the areas of robust and fast digital communication, video processing, and pattern recognition. These facilitate the development of effective and robust real-time systems.
- Digital video presents stringent throughput requirements for a multimedia communication system in terms of robustness and real-time performance.

Nevertheless, video information can be acquired, processed, and transmitted in different ways, and we have provided a panoramic view of such modalities in this issue.

Video communications aspects are fundamental in surveillance systems [54]–[59]. Data are acquired by distributed sources and then are usually transmitted to some remote control center. An important communication requirement is the bandwidth that should be lower for the down-link (from the control center to the sensors) than for the up-link (from the sensors to control center). Another important aspect is the security of the transmission. In many applications, surveillance data must be transmitted over open networks with multiuser access characteristics [18]. Information protection on such networks is a critical issue for maintaining privacy in the surveillance service. On the other hand, paternity of surveillance data can be very important for effective use for law enforcement purposes. Therefore, legal requirements necessitate the development of watermarking and data-hiding techniques for secure sensor identity assessment. Video processing and understanding requirements in surveillance systems are more severe than in classical computer vision systems due to the high variability and irregularity of the monitored scenes. Such variability has several consequences in required processing tools. From one point of view, it makes it necessary to use more sophisticated image processing algorithms for signal preprocessing and filtering. On the other hand, highly variable scene conditions imply the necessity of selecting robust scene description and pattern recognition methods. The automatic capability to learn and adapt to changing scene conditions and the learning of statistical models of normal event patterns are emerging issues in surveillance systems [42], [60]. The learning system provides a mechanism to flag potentially anomalous events by the discovery of the normal patterns of activity and flagging the least probable ones. Two major constraints that impact the deployment of these systems in

the real world include real-time performance and low cost [61]. Moreover, the multisensor aspect of a surveillance system constitutes a rather important direction for improving algorithms [2]. Multisensor systems can take advantage from processing either the same type of information acquired from different spatial locations or information acquired by sensors of different type (e.g., video cameras, microphones, etc.) on the same monitored area [3]. Appropriate processing techniques and new sensors providing the real-time information related to different scene characteristics can help both to enlarge the size of monitored environments and to improve performances of alarm detection in areas monitored by more sensors.

II. REVIEW OF THE STATE OF THE ART

Electronic video surveillance systems that have been proposed in literature can be classified under a technological perspective as belonging to three successive generations. The three generations follow the evolution of communications, processing, and storage and they have evolved in recent years with the same increasing speed of such technologies. Obviously, different categorizations can be established (see, e.g., [1]) that are based on different aspects of surveillance: for example, categories have been proposed to classify surveillance systems according to the degree of awareness of observed people being monitored. An excellent historical perspective is presented in [1] of the basic scientific discoveries that allowed surveillance video devices, storage media, and image transmission techniques to be progressively developed. Early breakthroughs in optics, including the discovery of lenses and concepts leading to the pinhole camera model, are shown to be as important as the more recent event understanding and recording tools (Daguerre [63]). The capability of observing and recording images from distant places has been originally oriented to monitor what happens in heaven. However, more prosaic observation of what happens on earth has been discovered by video-based surveillance to be as interesting; however, surveillance of events occurring on Earth poses ethical problems as such events often involve humans and the right to monitor can be in conflict with the individual privacy rights of the monitored people. These privacy problems largely depend on the shared acceptance of the surveillance task as a necessity by the public at large with respect to a given application. Another technological breakthrough fundamental to the development of surveillance systems is the capability of remotely transmitting and reproducing images and video information [e.g., TV broadcasting and the successive use of video signal transmission and display in close circuit TV systems (CCTV)]. CCTVs operative on the market and providing data at acceptable quality can be found dating back to 1960. The availability of CCTVs can be considered as the starting point that allowed on-line surveillance to be possible, and 1960 can be considered the starting date of the first generation surveillance systems.

First video surveillance systems (1GSS) (1960–80) basically extend human perception capabilities in a spatial sense. More “eyes” (i.e., video cameras) are used to display

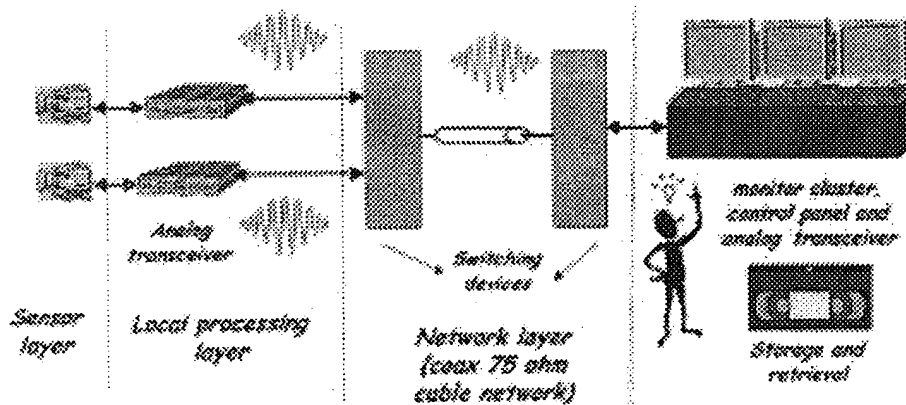


Fig. 1 Architectural example of first generation video-based surveillance system (1960–1980).

analog visual signals from multiple remote locations in a single physical location (i.e., the control room). 1GSSs are based on analog signal and image transmission and processing (Fig. 1). In these systems, video data from a set of cameras viewing remote scenes (sensor layer) are presented to the human operators after analog communication (local processing layer) of the video signal. Human operators analyzed video streams through a large set of monitors, where the scenes monitored by multiple cameras were multiplexed and presented in a periodic and predefined order. An added value of 1GSS is given by the acquired capability of telepresence of a human with respect to a remote place in a certain instant. Some major drawbacks of these systems have to do with the reasonably small attention span of operators that may result in a high miss rate of the events of interest. From a communications point of view, these systems suffered from the main problems of analog video communications: i.e., high bandwidth requirements, poor allocation flexibility, etc. Storage of video surveillance tapes remained a problem until the mid-1970s, when analog storage on VHS and similar media alleviated this problem.

The main limitations of the first generation systems are due to the following points strictly related to analog processing and transmission level.

- A large bandwidth is usually required that limits the number of sensors to be used [57].
- Analog video is subject to noise in transmission and the stored information suffers from degradations in image quality during playback [54]–[59].
- On-line alarm detection for a large set of monitored sites is difficult as they are related to visual inspection of monitors by human operators with limited attention spans [64].
- Off-line archival and retrieval of information on significant events of interest is difficult due to the large amount of tapes to be stored and reexamined.

It is clear from the above points that if either the spatial extent of the area being monitored or the complexity of events increases, then the only practical solution for real-time event detection using 1GSS is to increase the number of operators, i.e., to increase the number of parallel human processors for signals associated with events.

Starting from 1980, rapid improvements in the different basic technologies emerged: the improved resolution of video cameras and the availability of low-cost computers are two basic breakthroughs that facilitated intense research on algorithms for video processing and detection of events. In parallel, communications improvements during the 1980s led to CCTVs with improved robustness at reduced costs. In this technological evolution, *second generation surveillance systems* (2GSS) (1980–2000) correspond to the maturity phase of analog 1GSSs; they benefited from early advances in digital video communications (e.g., digital compression, bandwidth reduction, and robust transmission) and processing methods that provide assistance to the human operator by prescreening of important visual events. Some of these systems have been studied since the late 1980s until now in the context of different international research programs [65], [66] and have carried to prototypical products showing the feasibility of digital, intelligent attention focusing systems on video from limited sets of cameras.

In particular, 2GSS research addressed many areas with increased results in real-time analysis and segmentation of two-dimensional (2-D) image sequences [67], identification and tracking of multiple objects in complex scenes [68]–[73], human behavior understanding [35]–[45], multi-sensor data-fusion [74], intelligent man–machine interfaces [75]–[77], performance evaluation of video processing algorithms [78], [79], wireless and wired broad-band access networks [80]–[83], new signal processing for video compression, and multimedia transmission for video-based surveillance systems [84]–[90], etc.

Most research efforts during the period of 2GSSs have been spent on the development of automated real-time event detection techniques for video surveillance. As we have mentioned before, the availability of automated methods would greatly facilitate the monitoring of large sites with numerous cameras as the automated event detection step allows for pre-filtering and presentation of the relevant events.

In this way, the augmented perception capability in 2GSSs allows for a significant increase in the amount of simultaneously monitored data and, in addition, provides alarm data directly relevant to the cognitive monitoring tasks. Humans

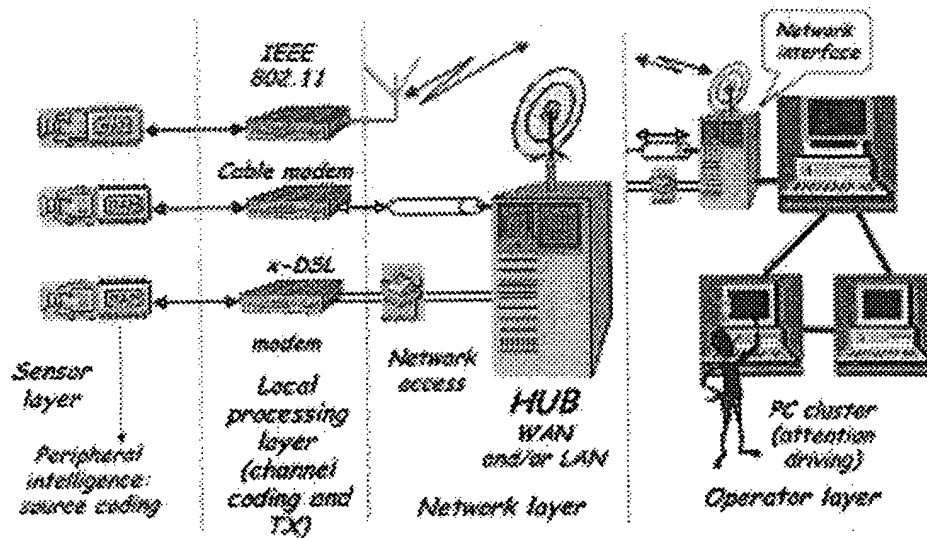


Fig. 2 Architectural example of third generation video-based surveillance system.

and animals are provided this ability through the use of preattentive mechanisms. It has been shown that there is evidence that neural nets implementing motion detection are used by the brain to capture human attention on specific sections of the human retina. Simple examples of multisensor extensions of this phenomenon are provided by the capability of humans to focus their sensors toward spatial areas from which specific sounds have been heard. However, 2GSSs have been able to only provide solutions with intermediate levels of digital video signal transmission and processing [80]–[90], i.e., they occasionally include digital methods in system subparts to solve local and isolated problems.

The main goal of *third generation surveillance systems* (3GSS) is to provide “full digital” solutions to the design of surveillance systems, starting at the sensor level, up to the presentation of mixed symbolic and visual information to the operators (see Fig. 2). In this sense, they take advantage of progress in low cost, high performance computing networks and in the availability of digital communications on heterogeneous, mobile, and fixed broad-band networks [56], [57].

In Fig. 2, an example of 3GSS is presented where video cameras constitute the sensor layer, while the peripheral intelligence and the transmission devices form the local processing layer. Sensor and local processing layers can be physically organized together in a so-called *intelligent camera*. The local processing layer uses digital compression methods to save bandwidth resources. The principal component of the network layer is the intelligent hub: the main functionality of the intelligent hub is the application-oriented fusion of data coming from lower-level layers. At the operator layer, an active interface is presented to the operator. This interface assists the operator by focusing his/her attention to a subset of interesting events. Communications are entirely in a digital form. The communication medium could be fixed wireless LANs or mobile digital devices (e.g., GPRS digital mobile phones) as well as broad-band media such as optical fibers, coax cables, or twisted pairs.

Research work on distributed real-time video processing techniques on intelligent, open, and dedicated networks is expected to provide more and more interesting results. This will be largely due to the availability of increased computational power at reasonable costs, advanced video processing/understanding methods, and multi-sensor data fusion. At the same time, a 3GSS can take advantage from the evolution of multimedia digital broadband communications in both wireless and wired domains. In particular, progress in the design of high-bandwidth access networks makes it possible to forecast widespread use of these systems by residential users for different applications. However, these surveillance systems would present specific requirements that necessitate the dedicated research and development of new tools.

This Special Issue is aimed at providing a global view of research efforts that are driving the development of 3GSSs as well as to provide an insight into the industrial perspectives of research centers developing them.

III. TECHNICAL CHALLENGES OF DEPLOYING THIRD GENERATION SYSTEMS

We have seen that the main objective of full digital 3GSSs is to facilitate the efficient data communication, management, and extraction of events in real-time video from a large collection of sensors. To achieve this goal, improvements in automatic recognition functionalities and digital multiuser communications strategies are needed. Technology meeting the requirements for the recognition algorithms includes computational speed, memory usage, remote data access, multiuser communications between distributed processors, etc. The availability of this technology greatly facilitates 3GSS development and deployment.

From the point of view of augmentation of human perception and monitoring capabilities in 3GSSs, the 3GSS alleviates the human from monitoring a collection of video monitors and, in addition, would assist the human in tasks that are

rather cumbersome (i.e., fall outside normal human spatial and temporal cognitive abilities) to do with traditional systems. For instance, real-time person tracking in a crowded scene is a tough task for a human to perform with a single video displayed on the monitor. Another improvement of 3GSSs is that online tools can be built to assist humans with event management.

A. Technological Viewpoint

If we consider technological aspects, one of the major technological basis of 3GSSs is the availability of robust, high-bandwidth digital multimedia transmissions over wide-band channels. Another technological basis has been the availability of embedded digital sensors that directly process locally acquired digital data. As progress in 3GSSs' intelligent sensors are being made, we are seeing the deployment of hubs capable of performing limited local digital video processing functions based on embedded DSPs. In addition, there is an increase in the amount of computing power per unit cost for use in the central control rooms and intelligent hubs, thus allowing automated intelligent processing to be done at the control center or in intermediate surveillance stations. Therefore, the driving technological push in 3GSSs is based on three main aspects.

- Wide-band digital communications and surveillance networking.
- Rapid decrease in processing hardware cost.
- Appearance of embedded intelligence subsystems (sensors and hubs).

Thanks to the availability of more evolved and powerful communications, sensors, and processing units, the architectural choice in 3GSSs can potentially become highly variable and flexibly customized to obtain a desired performance level. Therefore, the system architecture starts to represent a key issue; for example, the different level of distribution of intelligence can lead preattentive detection methods either closer to the sensors or distributed at different levels in a computational processing hierarchy. Another source of variability is due to the use of heterogeneous networks (wireless or wired) and transmission modalities both in terms of source and channel coding and in terms of multiuser access techniques. Spatial and temporal coding scalability can be very useful for reducing the amount of information to be transmitted by each camera depending on the intelligence level of the camera itself, while multiple access techniques are a basic tool to allow a large number of sensors to share a communication channel in the most efficient and robust way. Surveillance network management techniques are also necessary in 3GSSs to coordinate distributed intelligence modules in order to obtain a optimal performances as well as to adapt system behavior depending on the variety of conditions occurring either in a scene or in systems' parameters. All these tools are critical to design efficient systems. For example, the number of cameras supported by a system can vary to a large degree depending on both the level of intelligence embedded in each camera and on the channel capacity available for messages sent by cameras. Finally, a further evolution is

the integration among surveillance networks based on sensors of either different types such as audio, radar or always visual but oriented toward completely different functionalities (e.g., face detection, fingerprinting) and sensor types (e.g., standard perspective cameras or catadioptric sensors, i.e., sensors with mirrors).

The major technological improvements expected in 3GSSs can be structured onto different generality levels. This depends on the major complexity of these systems with respect to previous generations. Moreover, we can suppose that, due to such complexity, the development of a 3GSS system with all the characteristics underlined in the following cannot be reached until ten years from now. This also opens the problem of identifying successive progressive steps inside 3GSSs that can reasonably be integrated at successive stages into a single system.

Let us first analyze major improvements expected at different levels.

At a general level, a 3GSS should support:

- multiple services related to different users accessing to the same set of data acquired by a surveillance network (controlled multiuser accessibility);
- flexible changing of the functionalities assigned either to a single cell or to a group of cells depending on the active services as well as on operating conditions (cell reconfigurability).

A surveillance service should be complete and it should allow data accessibility both for direct alarm generation and for off-line inspection, i.e., it must include:

- a user oriented, sufficiently extended number of functionalities associated with a number of sensorial cells sufficient to provide a spatial surveillance support appropriate for the task (completeness);
- an alarm generation mechanism satisfying real-time alarm generation user requirements (real-time response);
- Distributed digital memorization capabilities and local databases accessible for a given time from the event and covering a sufficiently extended period (off-line recording).

Each functionality should be characterized by measurability, robustness, efficiency, multimodal sensor support, and adaptability with respect to both processing and communications. In particular, each functionality should be associated with the following.

- A computational model of a detection method appropriate to identify events of interest from available signal representations (computability).
- A measurable performance metric depending on the operating conditions (measurability).
- A performance behavior that should degrade gracefully with respect to the presence of various environmental conditions; such conditions should include the possibility for a functionality that can be applied to recorded, compressed data, by considering compression rate as an external condition (robustness).

Table 1
Real-World Applications

<i>Application Domain</i>	<i>Primary Benefits</i>	<i>Intelligent Functionality desired</i>	<i>Cost and Performance requirements</i>
Public area monitoring Large facility monitoring	Safety, Security	Person/vehicle detection, tracking and event analysis	Low system cost, false alarm/detection requirements rather stringent
Building exterior and interior monitoring, Parking Garage monitoring	Security, Safety, Access Control, Building Automation	Person/vehicle detection, parking space monitoring, license plate recognition, face recognition	High-end market High reliability desired in access control. Illumination is controlled / unconstrained
Subway, Highway, Tunnel monitoring, Transportation applications	Safety, Security, Resource management and Improved quality of service	People detection and tracking, vehicle, truck detection/tracking, classification of type of objects, recognition of events	Few high-end systems exist in the market, Very low false alarm rates. All weather and illumination conditions
Indoor monitoring (Malls, lobbies, Banks shopping complexes)	Security and Safety	Person detection, tracking, event analysis	Low cost systems, minimal false alarms

- A modifiable processing behavior to detect events of interest depending on environmental scene conditions (processing adaptability).
- A modifiable communication strategy depending on conditions of the channels (communications adaptability).
- A high ratio between performances with respect to employed computational and bandwidth resources (efficiency).
- An appropriate selection of sensors organized into system cells in order to provide data necessary to detect events of interest (multimodal sensorial support).

IV. RESEARCH IMPACT ON REAL-WORLD PRODUCT DEVELOPMENT

We have seen how the technological trends impact the research and development of the 3GSSs. The design, development, and deployment of these systems in the real world are influenced by a variety of factors including: the availability of sophisticated algorithms, the integration of the algorithms into the system form, and the validation that the system designed meets end user requirements. The industrial trends in CCTV systems are to incorporate intelligent processing functionality into these systems. High-end systems are being offered that take advantage of the broad-band communication capabilities and the intelligent algorithms available. However, their acceptance in the real world has been rather slow, mainly due to prohibitive cost of these systems and due to the end-user acceptance of these products (for a good discussion on end-user concerns and a market analysis of the security industry please see the paper from Pavlidis *et al.* of Honeywell Research in this Special Issue). Early use of CCTV

systems has been in large public installations (i.e., subway systems, large public areas) for improved safety and security, in military installations, private buildings, banks, and in shopping centers. More increasingly, video monitoring systems are being used in medium-scale shopping centers and in small shops. These are still based on 2GSS technology. Evaluation of a 2GSS system is primarily based on the quality of the image or video being presented to the user, on the number of video streams that can be monitored effectively. However, in a 3GSS system this is not the case. The intelligence functionality in a 3GSS system introduces the fundamental issue of validation of the intelligence component to verify that the alarm generation software meets user requirements. Since the end-users do not understand computer vision or signal processing technologies, their expectations for this technology are rather high at first glance. It is not uncommon for a highway authority official to expect people/vehicle detection and tracking error rates of less than 1% in all weather conditions, a task that is rather daunting even for humans. There is a need for 3GSS system researchers and designers to understand realistic use case scenarios of these systems and to translate end-user requirements to design practical and efficient systems. In Table I, we categorize real-world applications, their functional requirements, and cost/performance requirements.

The major application areas for 3GSSs are in the area of public monitoring. This is necessitated by rapid growth of metropolitan localities and by the growing need to provide improved safety and security to the general public. Other factors that drive the deployment of these systems include effective resource management, providing rapid emergency assistance, etc. The market for security and surveillance systems is slated to grow from about \$650 billion in the current year to about \$1.225 billion in the year 2006 worldwide.

Some factors that currently impede the deployment of these systems include:

- 1) system costs for given performance;
- 2) robustness of system functions with respect to complexity of input video (e.g., outdoor/natural illumination conditions, all weather conditions);
- 3) lack of standards for quantification of performance of these systems;
- 4) high costs, tediousness of tests and validation;
- 5) high-level vision functions providing semantics in video are rather error prone and generate too many false positives;
- 6) automated systems need to provide self-diagnosis when a scenario that is not modeled is encountered.

The system costs are rather prohibitive currently if one examines the level of performance required. Video surveillance is a visual task that is boring yet easy for a human operator to perform during short attention spans. End-users cannot comprehend the difficulty in the automation of such visual tasks. Due to their lack of understanding of vision systems, unrealistic performance requirements are often set. Nevertheless, the false alarm rates per camera for an event detection task should be rather low. This is driven by the psychological need for the human operator to trust the automated system. If the automated system generated too many false alerts, the human would tend to ignore the automated system and, hence, the intelligence function will be turned off. The problem is compounded when many types of events are automatically generated. The false alarms just add up. Typical system requirements for a people detection task in highways, for instance, is close to 100% detection with near zero false alarms per day under all weather conditions. A false alarm in this case is the detection of a change in the scene as a person. Another system requirement is the reaction time, i.e., the time it takes the system for an alarm to be generated, for these systems. Typical reaction times may vary depending on the event, but it is reasonable to expect reaction times of the order of a few seconds.

Another major stumbling block in incorporating these intelligence functions in real-world systems is the lack of robustness, the inability to test and validate these systems under variety of usage cases, and the lack of quantification of performance of these systems. A major requirement in automated systems is the ability to self-diagnose when the video data is not usable for analysis purposes. For instance, when CCD cameras are used in an outdoor highway application, it is often the case that during certain times of the day there is direct lighting of the camera lens from sunlight; a situation that renders the video useless for monitoring purposes. Another example of such a scenario is a weather condition such as heavy snowfall during which the contrast levels are such that people detection at a distance is rather difficult to do. Thus, in these scenarios, it is useful to have a system diagnostic that alerts the end-user of the unavailability of the automated intelligence functions. Ideally, the function that evaluates the unavailability of a given system should estimate whether the input data is such that the system performance can be guaranteed to meet given user-defined specifications. In addition,

the system should gracefully degrade in performance as the complexity of data increases. This is a very open research issue that is crucial to the deployment of these systems.

Performance evaluation of these systems, therefore, is a major open research issue. There is now a dedicated IEEE workshop on performance evaluation of tracking systems (PETS) that attempts to bring researchers to evaluate algorithms on common datasets to identify the algorithms strengths/limitations. However, there is a lack of realistic datasets and industrial input in these forums. Video databases that facilitate the systematic evaluation of the performance of various intelligent processing functions are needed. These databases should capture essentially all the variability in the scene conditions (e.g., day, night, day to night transitions, all object types, event types, dry, rainy, snow, foggy conditions) to effectively determine the situations under which the algorithms are effective and meet requirements. There is a need for performance metrics and well-agreed definitions for evaluating system components and the total system performance. Product development will benefit for the systematic comparisons of available methods. Testing and validation of these systems is rather costly and tedious due to the manual labor involved in validation. Intelligence functions can be built to have enough logged information to validate the alarms generated, while a periodic sampling/logging of the video data along with manual examination by a person is necessary to identify potentially missed alarms.

The first functionalities that we will see in the 3GSSs are intelligent detection and tracking functions with limited event analysis capabilities. Research systems currently have demonstrated these functionalities; see, for instance, [91]. The complexity of these event analysis methods is still rather low. They are primarily algorithms evaluating trajectories of movement patterns of people/vehicles to identify potential anomalies. The algorithms operate mainly in light pedestrian traffic conditions. More complicated event analysis functions will be needed to deal with moderate flow conditions. These will require multiple object tracking, reasoning, and interpretation of events.

V. SPECIAL ISSUE CONTENTS

In the previous sections, some of the main aspects were highlighted related to the current state of the art, technology, and industrial applications trends with respect to video surveillance systems. This Special Issue aims at providing a deeper insight in this topic by providing to the readers a balanced list of contributions of academic and industrial research aspects in communications, processing and understanding. As the reader will see from the papers of the Special Issue and as one can expect from the real-world problems explained in the previous section, main problems currently considered are related with real-time either distributed or centralized processing and robustness issues in multisensor surveillance networks. We hope that the invited papers presented by some of the more active research groups in this field will provide at the same time a sufficiently extended framework of current research status and new ideas for people who are interested in contributing to this interesting field where academic

approaches and industrial viewpoint can successfully meet to provide solutions from which real-world end-users can benefit. Nevertheless, we are also sure that this Special Issue necessarily covers only a limited part of the global work carried on in this field by not directly describing research of other academic and industrial groups in the world. Therefore, we invite interested readers to go through the references in various papers and in other Special Issues published in books and specialistic journals (e.g., [1]–[3], [61], [80], [90]–[92]) to enlarge the view we provided in this issue.

Referring to the contents of this special issue, we now present an overview of each of the invited and peer-reviewed published papers.

A. Change Detection and Background Extraction by Linear Algebra

(Invited Paper)

Durucan and Ebrahimi

The first paper in the Special Issue deals with a key issue in surveillance systems, i.e., optimal approaches to reduce the cardinality of data to be considered by further processing steps to obtain real-time scene descriptors. Change detection and background evaluation is particularly important in scenes observed by fixed cameras and can be managed in different ways depending by scene characteristics. In this first paper on change detection techniques as applied to video surveillance, the authors present an overview of several methods and discuss an innovative method that they have successfully applied in prototypical surveillance systems. The method is based on a physical luminance model and uses algebraic considerations to derive an estimation of the area of interest of an image with respect to an estimated background.

B. Into the Woods: Visual Surveillance of Noncooperative and Camouflaged Targets in Complex Outdoor Settings

(Invited Paper)

Boult, Micheals, Gao, and Eckmann

This paper discusses the current state of the art in video-based target detection with particular attention to the problem of surveillance and tracking of noncooperative and camouflaged targets in cluttered outdoor settings. Since for these domains, the detection phase is crucial, the authors discuss mainly techniques for change detection. Then, they present an innovative approach, called quasi-connected components (QCC), for performing spatio-temporal grouping. QCC combines gap filling, thresholding-with-hysteresis, and spatio-temporal region merging. The last part of the paper briefly review the tracking component of the system as well as the target geo-location, network communication, and user interface. Finally, the authors discuss the performance evaluation of the system, as measured by an external evaluation group.

C. Image Authentication Techniques for Surveillance Applications

(Invited Paper)

Bartolini, Tefas, Barni, and Pitas

The problem of image authentication in digital video surveillance systems is considered in this paper by authors

coming from two European universities very active in the watermarking research field. In particular, this paper provides an introductory overview to watermarking techniques where different approaches are discussed with their relative merits as compared to the considered application. This paper introduces the interesting viewpoint of designing watermarking algorithms in systems where quality is assessed not on the basis of a subjective/objective visual judgment but on the basis of indirect results i.e., automatic system decisions, like event detection in surveillance systems.

D. Distributed Architectures and Logical-Task Decomposition in Multimedia Surveillance Systems

(Invited Paper)

Marcenaro, Oberti, Foresti, and Regazzoni

Third generation video surveillance systems use distributed intelligence functionality. An important design issue is to decide the granularity at which the tasks can be distributed based on available computational resources, network bandwidth, and task requirements. The paper investigates the impact of distributed processing and communication techniques on the design of 3GSSs. The authors illustrate how the distribution of intelligence can be achieved by dynamic partition of all the logical processing tasks, including event recognition and communication. The dynamic task allocation problem is studied through the use of a computational complexity model for representation and communication tasks. The computational power of the intelligent cameras and the channel capacity of the bandwidth transmission are shown to be important parameters that affect the performance of the total system.

E. Multiple Camera Tracking of Interacting and Occluded Human Motion

(Invited Paper)

Dockstader and Tekalp

This paper describes a multicamera system for tracking interacting human motion based on multiple layers of temporal filtering coupled by a Bayesian belief network. The system uses a distributed platform, where a dedicated processor is applied to process each independent video stream representing a distinct view of some scene, to achieve real-time performance and to reduce overcome problems with occlusions and articulated motion. Each image of the monocular sequence is processed to extract interesting 2-D features (i.e., a set of image points to be tracked) of human motion. These measurements are used together with an estimate of the 3-D state vector representing the velocity and position of features in a 3-D Cartesian space as the input of a predictor-corrector filter that produces an estimate of the 2-D state vector. 2-D state vectors coming from each view of the system provide a vector of independent observations for a Bayesian belief network which fuses them to compute the most likely vector of 3-D state estimates given the available data. To maintain temporal continuity, the network is followed with a layer of Kalman filters that updates the 3-D state estimates. Experiments on a home environment with several people in motion demonstrate the superiority of

the proposed approach in tracking accuracy with respect to data fusion methods based on averaging.

F. Algorithms for Cooperative Multisensor Surveillance

(Invited Paper)

Collins, Lipton, Fuiyoshi, and Kanade

The Robotic Institute at Carnegie Mellon University (CMU) has created a Video Surveillance and Monitoring Lab. The team working in this laboratory has developed several video-understanding algorithms to perform cooperative, multisensor surveillance. An overview of these algorithms, integrated into a multicamera system, is described in this paper. The proposed system uses a distributed network of active video sensors to monitor activities in a cluttered outdoor environment. Video understanding algorithms are used to automatically detect people and vehicles, to localize and track them into a geo-spatial reference system and to classify them. Results from each single camera system are integrated into a coherent overview of the dynamic scene by multi-sensor fusion algorithms running on a central control room. Results are shown to a remote operator through a graphical user interface that provides the user with 2-D and 3-D synthetic views of the environment. Detected objects are displayed as dynamic agents.

The feasibility of the real-time functioning of the surveillance system has been demonstrated within a multicamera test-bed system developed on the CMU campus.

G. Urban Surveillance Systems: From the Laboratory to the Commercial World

(Invited Paper)

Pavlidis, Morellas, Tsiamyrtzis, and Harp

This paper describes a system developed in an industrial research center for the monitoring of a large building site parking lot with distributed set of cameras. The paper offers an industrial perspective to the security market as well as the end-user concerns. It discusses a system "DETER" that is used to detect and track pedestrians and vehicles in a parking lot using a distributed set of sensors. Various aspects of the system including background adaptation, object detection and tracking, trajectory analysis for threat identification, and visualization of the results from various sensors are presented. In addition, a qualitative as well as quantitative evaluation of the system performance is presented.

H. Design, Analysis, and Engineering of Video Monitoring Systems: An Approach and a Case Study

(Invited Paper)

Greifenhagen, Comaniciu, Niemann, and Ramesh

The problem of including a quantitative statistical performance evaluation model in the design of an industrial oriented multisensor surveillance system is the problem considered in this paper. The authors show first a general methodology by which a surveillance problem can be divided in a

set of submodules, each characterized by a precise statistical input-output relation. They show how performance of complex chains of such modules can be predicted in a statistical sense on the basis of probabilistic knowledge on input data. The industrial value of the paper is given by the case of study shown, where the problem of integrating data coming from an omni-directional camera to obtain an estimate of people position in an indoor scene can be used to point the optical axis of a different mobile camera toward the face of the observed people. The used performance evaluation model used in the design phase allows one to evaluate pointing error depending on the position of the observed people in the field of view of the omni-directional camera and to fix accordingly intrinsic and extrinsic parameters of the mobile camera to optimize the view of the observed face.

I. Aerial Video Surveillance and Exploitation

(Invited Paper)

Kumar, Sawhney, Samarasekera, Hsu, Tao, Guo, Hanna, Pope, Wildes, Hirvonen, Hansen, and Burt

This paper from an industrial research center describes a state-of-the-art aerial video surveillance system developed over several years of efforts for the Department of Defense Advanced Projects Agency in the U.S. The paper describes a framework for aerial video surveillance using video cameras. Aerial video surveillance is done delineating the video into components corresponding to static scene geometry, the dynamic objects, and the appearance of static/dynamic objects in the scene. The delineation is done based on 2-D/3-D alignment of dynamic imagery. Models that are progressively increasing complexity are invoked to delineate the static and dynamic components of the scene and efficiently represented for exploitation in various surveillance tasks. The paper discusses key components of the framework, including frame-to-frame alignment and the extraction of motion layers, mosaicing of static background components to form panoramas, independent tracking of moving objects, extraction of the geo-location of the video and tracked objects, and enhanced visualization of the video by reprojection and merging of the video with reference imagery and/or digital terrain maps. The system produces meta-data along with the video that allows one to perform aerial mapping, dynamic scene visualization over time, temporal change detection, etc.

CARLO S. REGAZZONI, *Guest Editor*
University of Genoa
Genoa I-16145, Italy

VISVANATHAN RAMESH, *Guest Editor*
Siemens Corporate Research Inc.
Princeton, NJ 08540 USA

GIAN LUCA FORESTI, *Guest Editor*
University of Udine
Udine 33100, Italy

ACKNOWLEDGMENT

The Guest Editors wish first to thank M. Kunt and the Proceedings Board for precious suggestions and generous encouragement to develop a framework proposal where to fit the original idea of this Special Issue. We wish also to thank J. Calder for his continuous and cooperative support during the period of preparation of this issue. Warmest thanks go to the invited authors who enthusiastically accepted our invitation of preparing high quality papers for this issue as well as to the following experts whose precious voluntary contribution as reviewers made it possible improvement of the quality of invited papers: T. Ellis, A. Venetsanopoulos, G. Thieil, O. Silven, Y. Kuno, T. Kalker, L. Cinque, F. Ramos, E. Memin, M. Tekalp, I. Pitas, T. Boult, D. Comaniciu, J. Llinas, V. Morellas, I. Pavlidis, I. Haritaoglu, J. Illingworth, T. Ebrahimi, N. Paragios, R. Kumar, D. Aubert, V. Roberto, J. Ferryman, F. Roli, M. Mustafa, P. Remagnino, S. Santini, G. Jones.

Moreover, the Guest Editors wish to thank F. Oberti and L. Marcenaro for their assistance in editorial activities.

REFERENCES

- [1] J. K. Petersen, *Understanding Surveillance Technologies*. Boca Raton, FL: CRC Press, 2001.
- [2] C. S. Regazzoni, G. Fabri, and G. Vernazza, *Advanced Video-Based Surveillance Systems*. Norwell, MA: Kluwer, 1998.
- [3] G. L. Foresti, P. Mahonen, and C. S. Regazzoni, *Multimedia Video-Based Surveillance Systems: Requirements, Issues and Solutions*. Norwell, MA: Kluwer, 2000.
- [4] H. Susama and M. Ukay, "Application of image processing for railways," *Q. RTRI*, vol. 30, no. 2, pp. 74-81, 1989.
- [5] U. Urfer, "Integration of systems and services in central monitoring stations (CMS)," in *Proc. IEEE Int. Carnahan Conf. Security Technology*, 1995, pp. 343-350.
- [6] M. Bogaert, N. Chelq, P. Cornez, C. S. Regazzoni, A. Teschioni, and M. Thonnat, "The PASSWORD project," in *Proc. Int. Conf. Image Processing*, Chicago, IL, 1996, pp. 675-678.
- [7] C. S. Regazzoni and A. Tesi, "Distributed data fusion for real-time crowding estimation," *Signal Process.*, vol. 53, pp. 47-63, 1996.
- [8] E. F. Lyon, "The application of automatic surface lights to improve airport safety," *IEEE AES Syst. Mag.*, pp. 14-20, 1993.
- [9] M. Braasch, M. DiBenedetto, S. Braasch, and R. Thomas, "LAAS operations in support of airport surface movement, guidance, control and surveillance: Initial test results," in *IEEE Proc. Position Location and Navigation Symp.*, 2000, pp. 82-89.
- [10] G. Galati, M. Ferri, P. Mariano, and F. Marti, "Advanced integrated architecture for airport ground movements surveillance," in *Radar Conf.*, 1995, pp. 282-287.
- [11] H. Wang, T. Kirubarajan, and Y. Bar-Shalom, "Precision large scale air traffic surveillance using IMM/assignment estimators," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, pp. 255-266, Jan. 1999.
- [12] B. Sridhar and G. B. Chatterji, "Computationally efficient conflict detection methods for air traffic management," in *Proc. American Control Conf.*, vol. 2, 1997, pp. 1126-1130.
- [13] G. Donohue, "Vision on aviation surveillance systems," in *Proc. IEEE Int. Radar Conf.*, 1995, pp. 1-4.
- [14] G. L. Foresti and B. Pani, "Monitoring motorway infrastructures for detection of dangerous events," in *IEEE Proc. Int. Conf. Image Analysis and Processing*, Venice, Italy, 1999, pp. 1144-1147.
- [15] J. M. Manendez, L. Salgado, E. Rendon, and N. Garcia, "Motorway surveillance through stereo computer vision," in *IEEE Proc. 33rd Annu. Int. Carnahan Conf. Security Technology*, 1999, pp. 197-202.
- [16] D. Koller, K. Daniilidis, and H. Nagel, "Model-Based object tracking in monocular image sequences of road traffic scenes," *Int. J. Comput. Vis.*, vol. 10, pp. 257-281, 1993.
- [17] J. Malik, D. Koller, and J. Weber, "Robust multiple car tracking with occlusion reasoning," in *Eur. Conf. Computer Vision*, Stockholm, Sweden, 1994, pp. 189-196.
- [18] S. H. Park, K. Jung, J. K. Hea, and H. J. Kim, "Vision-Based traffic surveillance system on the internet," in *Proc. 3rd Int. Conf. Computational Intelligence and Multimedia Applications (ICCIMA '99)*, 1999, pp. 201-205.
- [19] J. E. Boyd, J. Meloche, and Y. Vardi, "Statistical tracking in video traffic surveillance," in *Proc. 7th IEEE Int. Conf. Computer Vision*, vol. 1, 1999, pp. 163-168.
- [20] A. F. Toal and H. Buxton, "Spatio-temporal reasoning within a traffic surveillance system," in *Proc. 2nd Eur. Conf. Computer Vision*, S. Margherita, Ed., Italy, 1992, pp. 884-892.
- [21] J. M. Blasseville, "Image processing for traffic management," in *Advanced Video-Based Surveillance Systems*, C. S. Regazzoni, G. Vernazza, and G. Fabri, Eds. Norwell, MA: Kluwer, 1999, pp. 67-75.
- [22] C. P. K. Sherwood, "Traffic surveillance and control systems for the area of Hong Kong," in *9th Int. Conf. Road Transport Information and Control*, Apr. 21-23, 1998, pp. 191-194.
- [23] S. A. Hamid, S. A. Rahman, and J. J. Steed, "The introduction of traffic surveillance and control on the privatised express ways in Malaysia," in *9th Int. Conf. Road Transport Information and Control*, Apr. 21-23, 1998, pp. 200-206.
- [24] R. B. Olsen, P. Bugden, Y. Andrade, P. Hoyt, M. Lewis, H. Edel, and C. Bjerkelund, "Operational use of RADARSAT SAR for marine monitoring and surveillance," in *Int. Symp. Geoscience and Remote Sensing. IGARSS '95*, vol. 1, 1995, pp. 224-226.
- [25] A. N. Ince and E. Topuz, "The design and computer simulation of a maritime surveillance system," in *Int. Conf. Radar*, 1997, pp. 653-656.
- [26] K. Takasaki, T. Sugimura, and S. Tanaka, "Comparison of sea traffic in Tokyo and Osaka Bays with JERS-1/OPS data," in *Proc. Int. Geoscience and Remote Sensing Symp., IGARSS '96*, pp. 79-81.
- [27] J. G. Sanderson, M. K. Teal, and T. J. Ellis, "Target identification in complex maritime scenes," in *6th Int. Conf. Image Processing and its Applications*, vol. 2, 1997, pp. 463-467.
- [28] C. A. Rodriguez, J. A. Howell, H. O. Menlove, C. M. Brislawn, J. N. Bradley, P. Chare, and T. Gorten, "NUCLEAR video image processing for nuclear safeguards," in *Proc. IEEE 29th Annu. Int. Carnahan Conf. Security Technology*, 1995, pp. 355-363.
- [29] B. B. Berson, R. S. Wallace, and E. L. Schwartz, "A miniaturized active vision system," in *Proc. 2nd Int. Conf. Pattern Recognition*, vol. 4, 1992, pp. 58-61.
- [30] M. J. Cattle, "The use of digital CCTV in an airport car-park application," in *Proc. IEEE 29th Annu. Int. Carnahan Conf. Security Technology*, 1995, pp. 180-185.
- [31] E. Stringa and C. S. Regazzoni, "Content-based retrieval and real time detection from video sequences acquired by surveillance systems," in *IEEE Int. Conf. Image Processing*, vol. 3, Chicago, IL, Oct. 4-7, 1998, pp. 138-142.
- [32] C. Lin and R. Nevatia, "Building detection and description from a single intensity image," *Comput. Vis. Image Understand.*, vol. 72, no. 2, pp. 101-121, 1998.
- [33] L. Vergara and P. Bernabeu, "Automatic signal detection applied to fire control by infrared digital signal processing," *Signal Process.*, vol. 80, no. 4, pp. 659-669, 2000.
- [34] W. Millesi, M. J. Truppe, F. Watzinger, A. Wagner, and R. Ewers, "Image guided surgery extended by remote stereotactic visualization," in *Lecture Notes in Computer Science*, J. Troccaz, E. Grimson, and R. Mosges, Eds. Berlin, Germany: Springer, 1997, vol. 1205, pp. 813-821.
- [35] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 809-830, Aug. 2000.
- [36] J. Heikkila and O. Silven, "A real-time system for monitoring of cyclists and pedestrians," in *Proc. 2nd IEEE Workshop on Visual Surveillance (VS'99)*, 1999, pp. 74-81.
- [37] S. Hongeng, F. Bremond, and R. Nevatia, "Representation and optimal recognition of human activities," in *Proc. Int. Conf. Computer Vision and Pattern Recognition (CVPR2000)*, 2000, pp. 818-825.
- [38] D. M. Gavrilu, "The analysis of human movement and its application for visual surveillance," in *Proc. 2nd IEEE Workshop on Visual Surveillance (VS'99)*, 1999.
- [39] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 831-843, 2000.
- [40] Q. Cai and J. K. Aggarwal, "Tracking human motion using multiple cameras," in *Proc. 13th Int. Conf. Pattern Recognition*, Vienna, Austria, Aug. 25-29, 1996, pp. 68-72.

- [41] Y. Ricquebourg and P. Bouthemy, "Real-time tracking of moving persons by exploring spatio-temporal image slices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 797–808, 2000.
- [42] A. Galata, N. Johnson, and D. Hogg, "Learning behavior models of human activities," in *Proc. British Machine Vision Conf.*, U.K., 1999.
- [43] J. Aranda, J. Amat, and M. Fragola, "A multitasking system for trajectory analysis of people in a restricted area," in *Proc. 4th Int. Workshop on Time-Varying Image Processing and Moving Object Recognition*, Florence, Italy, 1993.
- [44] J. A. Freer, B. J. Beggs, H. L. Fernandez-Canque, F. Chevrier, and A. Goryashko, "Automatic intruder detection incorporating intelligent scene monitoring with video surveillance," in *Proc. Eur. Conf. Security and Detection, ECOS 97*, 1997, pp. 109–113.
- [45] F. Bremond and M. Thonnat, "Tracking multiple nonrigid objects in video sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 585–591, 1998.
- [46] H. Buxton and S. Gong, "Visual surveillance in a dynamic and uncertain world," *Artif. Intell.*, vol. 78, no. 1–2, pp. 431–459, 1995.
- [47] C. Benabdelkader, P. Burlina, and L. Davis, "Single camera multiplexing for multi-target tracking," in *Multimedia Videobased Surveillance Systems. Requirements, Issues and Solution*, C. S. Regazzoni, G. L. Foresti, and P. Mahonen, Eds. Norwell, MA: Kluwer, 2000, pp. 130–142.
- [48] D. A. Prichard, "System overview and applications of a panoramic imaging perimeter sensor," in *Proc. IEEE 29th Annu. Int. Carnahan Conf. Security Technology*, 1995, pp. 420–425.
- [49] U. Oppelt, "New possibilities for video applications in the security field," in *Proc. IEEE 29th Annu. Int. Carnahan Conf. Security Technology*, 1995, pp. 426–435.
- [50] B. Peters, J. Meehan, D. Miller, and D. Moore, "Sensor link protocol: Linking sensor systems to the digital battlefield," in *Proc. IEEE Military Communications Conf.*, vol. 3, 1998, pp. 919–923.
- [51] M. T. Fennell and R. P. Wishner, "Battlefield awareness via synergistic SAR and MTI exploitation," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 13, pp. 39–43, Feb. 1998.
- [52] J. S. Draper, S. Perlman, C. K. Chuang, M. Hanson, L. Lillard, B. Hibbeln, and D. Sene, "Tracking and identification of distant missiles by remote sounding," in *Proc. IEEE Aerospace Conf.*, vol. 4, 1999, pp. 333–341.
- [53] G. A. V. Sickie, "Aircraft self reports for military air surveillance," in *Proc. IEEE Digital Avionics Systems Conf.*, vol. 2, 1999, pp. 2–8.
- [54] C. S. Regazzoni, C. Sacchi, and C. Dambra, "Remote cable-based video surveillance applications: the AVS-RIO project," in *Proc. ICLAP99*, Venice, Italy, Sept. 27–29, 1999, pp. 1214–1215.
- [55] F. Soldatini, P. Mahonen, M. Saarinen, and C. S. Regazzoni, "Network management within an architecture for distributed hierarchical digital surveillance systems," in *Multimedia Videobased Surveillance Systems. Requirements, Issues and Solutions*, C. Regazzoni, G. Foresti, and P. Mahonen, Eds. Norwell, MA: Kluwer, pp. 143–157.
- [56] P. Mahonen and M. Saarinen, "Broadband multimedia transmission for surveillance applications," in *Multimedia Videobased Surveillance Systems. Requirements, Issues and Solutions*, C. S. Regazzoni, G. L. Foresti, and P. Mahonen, Eds. Norwell, MA: Kluwer, pp. 173–185.
- [57] K. Pahlavan and A. H. Levesque, "Wireless data communications," *Proc. IEEE*, vol. 82, no. 9, pp. 1398–1430, 1994.
- [58] —, *Wireless Information Networks*. New York: Wiley, 1995.
- [59] S. Glisic and B. Vucetic, *Spread Spectrum CDMA Systems for Wireless Communications*. Norwood, MA: Artech House, 1997.
- [60] M. Walter, A. Psarrou, and S. Gong, "Learning prior knowledge and observation augmented density models for human behavior recognition," in *Proc. British Machine Vision Conf.*, U.K., 1999.
- [61] G. L. Foresti and C. S. Regazzoni, "Video processing and communications in real-time surveillance systems," *J. Real-Time Imaging*, vol. 7, no. 3, 2001.
- [62] G. Gernsheim, H. Helmut, A. Allison, and L. J. M. Daguerre, *The History of the Diorama and the Daguerreotype*. New York: Dover, 1968.
- [63] L. J. M. Daguerre, "Histoire et description des procedes du daguerreotype et du diorama Daguerre's Manual," 1839.
- [64] C. H. M. Donald, "Assessing the human vigilance capacity of control room operators," in *Proc. Int. Conf. Humans Interfaces in Control Rooms, Cockpits and Command Centres*, 1999, pp. 7–11.
- [65] ESPRIT Program, European Union. [Online]. Available: <http://www.newcastle.research.ec.org/esp-syn/all-ac-index.html>
- [66] VSAM Program, USA. [Online]. Available: <http://www.cs.cmu.edu>
- [67] S. M. Smith, "ASSET-2: Real-time motion segmentation and object tracking," *Real Time Imaging*, vol. 4, pp. 21–40, 1998.
- [68] G. L. Foresti, "Object detection and tracking in time-varying and badly illuminated outdoor environments," *Opt. Eng.*, vol. 37, no. 9, pp. 2550–2564, 1998.
- [69] Z. Li and H. Wang, "Real-time 3-D motion tracking with known geometric models," *Real Time Imaging*, vol. 5, pp. 167–187, 1999.
- [70] P. J. L. V. Beek, A. M. Tekalp, N. Zhuang, I. Celasun, and M. Xia, "Hierarchical 2-D mesh representation, tracking and compression for object-based video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 617–634, 1999.
- [71] D. B. Gennery, "Visual tracking of known 3D objects," *Int. J. Comput. Vis.*, vol. 7, no. 3, pp. 243–270, 1992.
- [72] D. G. Lowe, "Robust model-based motion tracking through the integration of searching and estimation," *Int. J. Comput. Vis.*, vol. 8, no. 2, pp. 113–122, 1992.
- [73] F. Meyer and P. Bouthemy, "Region-based tracking using affine motion models in long image sequences," *Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 60, no. 2, pp. 119–140, 1994.
- [74] P. K. Varshney, "Multisensor data fusion," *Electron. Commun. Eng. J.*, vol. 9, no. 6, pp. 245–253, 1997.
- [75] A. Pentland, "Looking at people: Sensing for ubiquitous and wearable computing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 107–119, 2000.
- [76] D. S. Faulstich and R. T. Ng, "An expressive language and interface for image querying," *Mach. Vis. Applicat.*, vol. 10, no. 2, pp. 74–85, 1997.
- [77] D. P. Haanpaa and G. P. Roston, "An advanced haptic system for improving man-machine interfaces," *Comput. Graphics*, vol. 21, no. 4, pp. 443–449, 1997.
- [78] F. Overti, E. Stringa, and G. Vernazza, "Performance evaluation criterion for characterizing video surveillance systems," *Real-Time Imaging J.*, vol. 7, no. 3, 2001.
- [79] T. Kanungo, M. Y. Jaisimha, J. Palmer, and R. M. Haralick, "A methodology for quantitative performance evaluation of detection algorithms," *IEEE Trans. Image Processing*, vol. 4, no. 12, pp. 1667–1673, 1995.
- [80] W. W. Lu, "Special issue on multidimensional broadband wireless technologies and services," *IEEE Trans. Commun.*, vol. 89, Jan. 2001.
- [81] K. Fazel, P. Robertson, O. Klank, and F. Vanselow, "Concept of a wireless indoor video communications system," *Signal Process. Image Commun.*, vol. 12, no. 2, pp. 193–208, 1998.
- [82] P. Scotton, "Compression and transmission of video signals over high speed networks with rate based congestion control," *Signal Process.*, vol. 36, no. 3, pp. 392–392, 1994.
- [83] P. Batra and S. F. Chang, "Effective algorithms for video transmission over wireless channels," *Signal Process. Image Commun.*, vol. 12, no. 2, pp. 147–166, 1998.
- [84] B. S. Manjunath, T. Huang, A. M. Tekalp, and H. J. Zhang, "Introduction to the special issue on image and video processing for digital libraries," *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 1–2, 2000.
- [85] E. Stringa and C. S. Regazzoni, "Real-time video-shot detection for scene surveillance applications," *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 69–79, 2000.
- [86] G. Bjontegaard, K. O. Lillevold, and R. Danielsen, "A comparison of different coding formats for digital coding of video using MPEG-2," *IEEE Trans. Image Processing*, vol. 5, no. 8, pp. 1271–1276, 1996.
- [87] H. Cheng and X. Li, "Partial encryption of compressed images and videos," *IEEE Trans. Signal Processing*, vol. 48, no. 8, pp. 2439–2451, 2000.
- [88] J. Benoispineau, F. Morier, D. Barba, and H. Sanson, "Hierarchical segmentation of video sequences for content manipulation and adaptive coding," *Signal Processing*, vol. 66, no. 2, pp. 181–201, 1998.
- [89] N. Vasconcelos and A. Lippman, "Statistical models of video structure for content analysis and characterization," *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 3–19, 2000.
- [90] T. Ebrahimi and P. Salembier, "Special issue on video sequence segmentation for content-based processing and manipulation," *Signal Processing*, vol. 66, no. 2, pp. 123–124, 1998.
- [91] R. Collins, A. Lipton, and T. K. Kanade, "Special issue on video surveillance and monitoring," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 745–746, 2000.
- [92] S. Maybank and T. Tan, "Introduction—Surveillance," *Int. J. Comput. Vis.*, vol. 37, no. 2, pp. 173–173, 2000.



Carlo Regazzoni (Senior Member, IEEE) was born in Savona, Italy, in 1963. He received the Laurea degree in electronic engineering and the Ph.D. degree in telecommunications and signal processing from the University of Genoa, Italy, in 1987 and 1992, respectively.

Since 1998, he has been Professor of Telecommunication Systems in the Engineering Faculty of the University of Genova. Since 1998, he has been responsible for the Signal Processing and Telecommunications (SP&T) Research Group at the Department of Biophysical and Electronic Engineering (DIBE), University of Genova, that he joined in 1987. His main current research interests are multimedia and nonlinear signal and video processing, signal processing for telecommunications, multimedia broad-band wireless, and wired telecommunications systems. He has been involved in research on multimedia surveillance systems since 1988. He has been co-organizer and chairman of the first two International Workshops on Advanced Video Based Surveillance, held in Genova, Italy, 1998 and Kingston, U.K., 2001. He has also organized several Special Sessions in the same field at International Conferences [Image Analysis and Processing, Venice 1999 (ICIAP99), European Signal Processing Conf. (Eusipco2000), Tampere Finland, 2000]. He has been responsible for several EU research and development projects dealing with video surveillance methodologies and applications in the transport field (ESPRIT Dimus, Athena, Passwords, AVS-PV, AVS-RIO). He has been also responsible for several research contracts with Italian industries; he served as a referee for international journals and as reviewer for EU in different research programs. He is a consultant for the EU Commission for the definition of the 6th research framework program in the Ambient Intelligence domain. He is co-editor of the books *Advanced Video-based Surveillance Systems* (1999) and *Multimedia Surveillance Systems* (2000). He is author or co-author of 43 papers in international scientific journals and of more than 130 papers presented at refereed international conferences.

Dr. Regazzoni is a Member of AEI and IAPR



Visvanathan Ramesh received the B.S. degree in engineering from the College of Engineering, Guindy, Madras, India in 1984, the M.S. degree in electrical engineering from Virginia Tech in 1987, and the doctoral degree from the University of Washington, where he defended his Ph.D dissertation titled "Performance characterization of image understanding algorithms" in 1994.

He is currently a Senior Member of Technical Staff and Project Manager of the real-time imaging effort in the Imaging Department at Siemens Corporate Research, Princeton, NJ. At Siemens, he has focused on the research and development of statistical methods for real-time video analysis functions such as object detection, tracking and action recognition. His most recent work is focused on the research of how contextual models and domain knowledge influence the selection of algorithms and its parameters for a given video analysis application. He supervises Ph.D students in several universities, including CMU, UIUC, Lehigh and U of Rochester. He also has advised several M.S. students from various universities in Europe. He

has been actively involved in image and video understanding research in low- and mid-level vision over the past 12 years and has published numerous publications on the topic. His primary objective is to build robust image and video analysis systems and to quantify robustness of IU algorithms. During the course of his Ph.D. work, he developed a systems engineering methodology for computer vision algorithm performance characterization and design. He has also focused on the development of software environments for computer vision. Besides his deep involvement in the Unix version of GIPSY (a general image processing system), he was a member of the ARPA Image Understanding Environment Committee, the committee that designed the IUE (an object oriented environment for Image Understanding Research). He was also part of a team that helped design the Java Advanced Imaging Specification (Sun Microsystem's Java API for advanced imaging). He has published several papers in the computer vision field, with large emphasis in the area of performance analysis of vision systems. He is a co-author of a paper on real-time tracking that got the best paper award in CVPR 2000. His broad research interests are pattern recognition, computer vision, artificial intelligence, and biomedical engineering.



Gian Luca Foresti (Senior Member, IEEE) was born in Savona, Italy, in 1965. He received the Laurea degree (*cum laude*) in electronic engineering and the Ph.D. degree in computer science from the University of Genoa, Italy, in 1990 and in 1994, respectively.

In 1994, he was visiting Professor at University of Trento, Italy. Since 1998, he has been Professor of the Computer Science at the Department of Mathematics and Computer Science (DIMI), University of Udine. His main interests involve artificial neural networks, multisensor data fusion, computer vision and image processing, and multimedia databases. Techniques proposed found applications in the following fields: automatic video-based systems for surveillance and monitoring of outdoor environments (e.g., underground stations, railway lines, motorway, etc.), vision systems for autonomous vehicle driving and/or road traffic control, 3-D scene interpretation, and reconstruction. He is author or co-author of more than 100 papers published in International Journals and Refereed International Conferences. He was general co-chair, chairman and member of Technical Committees at several conferences. He has been

co-organizer of three Special Sessions on video-based surveillance systems at International Conferences (ISATA97, ISATA98, ICIAP99). He has contributed to five books in his area of interest, and he is co-author of the book *Multimedia Systems for Visual-Based Surveillance* (Kluwer, 2000).

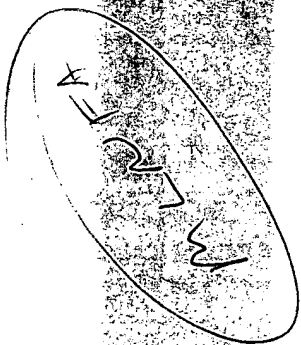
Dr. Foresti was Guest Editor of the Special Issue of *Real Time Imaging Journal* on "Video Processing and Communications in Real Time Video-Based Surveillance" and recently he was Guest Editor of a Special Issue of the PROCEEDINGS OF THE IEEE on "Video Communications, Processing and Understanding for Third Generation Surveillance Systems." He was an invited speaker at the NATO School on Multisensor Data Fusion, at Pitlocry, U.K., July 2000. He has served as a reviewer for several international journals and for the European Union in different research programs (MAST III, Long Term Research, Brite-CRAFT). He has been responsible for DIMI for several European and national research projects in the field of video-based surveillance for unattended outdoor environments. In February 2000, he was appointed as an Italian Member of the Information Systems Technology (IST) panel of the NATO-RTO. He is a Senior Member of LAPR.

AN EQUAL OPPORTUNITY EMPLOYER

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, \$300



NOT DELIVERABLE
AS ADDRESSED
UNABLE TO FORWARD



RECEIVED
DEC 27 2006
USPTO MAIL CENTER